Worley, C.
10/691590
Seq IDS 1 & 3

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 5, 2005, 00:24:30 ; Search time 4384.22 Seconds
(without alignments)
983.645 Million cell updates/sec

Title: US-10-691-590-1
Perfect score: 458
Sequence: 1 RREQEEESEEETFGEFQQVK.....FLNHKQNTNVIKFTVKASAY 89

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10691590/runat_02112005_120021_13145/app_query.fasta_1.462
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10691590_@CGN_1_1_3851_@runat_02112005_120021_13145 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	458	100.0	3318	8	TCCSV	X62625 T.cacao csv
2	339	74.0	1867	6	A20606	A20606 67kD protei
3	190	41.5	1248	8	LCU551425	AJ551425 Lens culi
4	185	40.4	1248	8	AJ626897	AJ626897 Pisum sat
5	185	40.4	1248	8	AJ626898	AJ626898 Pisum sat
6	185	40.4	1596	8	PSVICK	X67429 Pisum sativ
7	183	40.0	907	8	PEAVIC2	J01258 pea vicilin
8	182.5	39.8	1732	8	AY102931	AY102931 Juglans n
9	181	39.5	114452	8	AC148289	AC148289 Medicago
10	177.5	38.8	2647	8	AB019195	AB019195 Cucurbita
11	177	38.6	1257	8	LCU551424	AJ551424 Lens culi
12	175	38.2	1690	8	LCVICNC	Z48434 L.culinaris
13	175	38.2	2777	8	COTSPA	M19378 G.hirsutum
14	175	38.2	7876	12	SYNAGABA	M18027 Synthetic G
15	174.5	38.1	2057	8	AF066055	AF066055 Juglans r
16	174	38.0	1743	8	LEVICN	Z48438 L.ervoides
17	172.5	37.7	1693	8	LCOVICN	Z48437 L.culinaris
18	170	37.1	1741	8	LCVICNE	Z48440 L.culinaris
19	169	36.9	1561	8	VFVIC	Y00462 Vicia faba
20	169	36.9	4239	8	PSVICIL	X14076 Pea vicilin
21	168.5	36.8	5327	8	VFVICG	Y00506 Vicia faba
22	168	36.7	1742	8	LCVICNA	Z48436 L.culinaris
23	166	36.2	1433	8	PSVIC	Y00722 Pisum sativ
24	165	36.0	1690	8	LNVICN	Z48439 L.nigricans
25	164	35.8	1548	8	VNVICLN	Z71987 V.narbonens
26	164	35.8	1965	8	COTSPC	M16891 G.hirsutum
27	161	35.2	3736	8	PSCVCA	X06398 Pisum sativ
28	159	34.7	1858	8	VNCONVN	Z71986 V.narbonens
29	157.5	34.4	314	8	LES270964	AJ270964 Lycopersi
30	156.5	34.2	1418	8	AY581851	AY581851 Arachis h
31	156.5	34.2	1930	6	BD172109	BD172109 Peanut al
32	156.5	34.2	1930	6	AR257463	AR257463 Sequence
33	156.5	34.2	1930	6	BD107898	BD107898 Methods a
34	156.5	34.2	2032	6	AR257469	AR257469 Sequence
35	156.5	34.2	2032	6	AX148742	AX148742 Sequence
36	156.5	34.2	2032	6	AX155331	AX155331 Sequence
37	156.5	34.2	2032	8	ARQARAHI	L34402 Arachis hyp
38	156.5	34.2	2041	6	AR257470	AR257470 Sequence
39	156.5	34.2	2674	8	AF432231	AF432231 Arachis h
40	156	34.1	2069	8	PSA276875	AJ276875 Pisum sat
41	156	34.1	2212	8	LCU276683	AJ276683 Lens culi
42	154.5	33.7	1978	6	AX155332	AX155332 Sequence
43	153.5	33.5	1035	8	AY581850	AY581850 Arachis h
44	152.5	33.3	1595	8	OSU45322	U45322 Oryza sativ
45	151.5	33.1	3636	8	SOYBPSP	M13759 Glycine max

ALIGNMENTS

RESULT 1
TCCSV
LOCUS TCCSV 3318 bp DNA linear PLN 03-DEC-1993
DEFINITION T.cacao csv gene for seed vicilin.
ACCESSION X62625 S38078
VERSION X62625.1 GI:21910
KEYWORDS csv gene; seed protein; vicilin.
SOURCE Theobroma cacao (cacao)
ORGANISM Theobroma cacao
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae; Theobroma.
REFERENCE 1 (bases 1 to 3318)
AUTHORS McHenry,L. and Fritz,P.J.
TITLE Comparison of the structure and nucleotide sequences of vicilin genes of cocoa and cotton raise questions about vicilin evolution
JOURNAL Plant Mol. Biol. 18 (6), 1173-1176 (1992)
MEDLINE 92288309
PUBMED 1600151
REFERENCE 2 (bases 1 to 3318)
AUTHORS McHenry,L.
TITLE Direct Submission
JOURNAL Submitted (15-OCT-1991) L. McHenry, Pennsylvania State University, 111 Borland Lab, University Park, PA 16802, USA
COMMENT See also x62626
Overlap of sequenced fragments.
FEATURES Location/Qualifiers

```
     source          1. .3318
                     /organism="Theobroma cacao"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3641"
                     /clone="pHD5P1.7, pHD5P1.2, pHD5E4.5"
                     /tissue_type="leaves"
                     /clone_lib="cocoa genomic library in lambda GEM II"
                     /dev_stage="mature tree"
     TATA_signal     543. .549
                     /note="csv gene"
     gene            610. .2581
                     /gene="csv"
     CDS             join(610. .1240,1331. .1506,1597. .1677,1780. .2064,
                     2177. .2581)
                     /gene="csv"
                     /codon_start=1
                     /product="vicilin"
                     /protein_id="CAA44493.1"
                     /db_xref="GI:21911"
                     /db_xref="GOA:Q43358"
                     /db_xref="UniProt/Swiss-Prot:Q43358"
                     /translation="MVISKSPFIVLIFSLLLSFALLCSGVSAYGRKQYERDPRQQYEQ
                     CQRRCESEATEEREQEQCEQRCEREYKEQQRQQEEELQRQYQQCQGRCQEQQQGQREQ
                     QQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQRNNPYYFPKRRSFQTRFRDEE
                     GNFKILQRFAENSPPLKGINDYRLAMFEANPNTFILPHHCDAEAIYFVTNGKGTITFV
                     THENKESYNVQRGTVVSVPAGSTVYVVSQDNQEKLTIAVLALPVNSPGKYELFFPAGN
                     NKPESYYGAFSYEVLETVFNTQREKLEEILEEQRGQKRQQGQQGMFRKAKPEQIRAIS
                     QQATSPRHRGGERLAINLLSQSPVYSNQNGRFFEACPEDFSQFQNMDVAVSAFKLNQG
                     AIFVPHYNSKATFVVFVTDGYGYAQMACPHLSRQSQGSQSGRQDRREQEEESEEETFG
                     EFQQVKAPLSPGDVFVAPAGHAVTFFASKDQPLNAVAFGLNAQNNQRIFLAGRPFFLN
                     HKQNTNVIKFTVKASAY"
     sig_peptide     610. .681
                     /gene="csv"
     mat_peptide     join(682. .1240,1331. .1506,1597. .1677,1780. .2064,
                     2177. .2578)
                     /gene="csv"
                     /product="vicilin"
     exon            <610. .1240
                     /gene="csv"
                     /number=1
     intron          1241. .1330
                     /gene="csv"
                     /number=1
     exon            1331. .1506
                     /gene="csv"
                     /number=2
     intron          1507. .1596
                     /gene="csv"
                     /number=2
     exon            1597. .1677
                     /gene="csv"
                     /number=3
     intron          1678. .1779
                     /gene="csv"
                     /number=3
     exon            1780. .2064
                     /gene="csv"
                     /number=4
     intron          2065. .2176
                     /gene="csv"
                     /number=4
     exon            2177. .>2582
                     /number=5
     polyA_signal    2856. .2861
ORIGIN

Alignment Scores:
Pred. No.:              2.88e-46        Length:         3318
Score:                  458.00          Matches:        89
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0
```

```
US-10-691-590-1 (1-89) x TCCSV (1-3318)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2312 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 2371

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2372 GCCCCATTGTCACCTGGTGACGTCTTTGTAGCCCCGGCAGGCCATGCAGTTACATTCTTT 2431

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2432 GCATCCAAAGACCAGCCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAGAACAACCAG 2491

Qy         61 ArgIlePheLeuAlaGlyArgProPhePheLeuAsnHisLysGlnAsnThrAsnValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2492 AGAATTTTCCTTGCAGGTAGGCCTTTTTTTCTGAATCACAAGCAAAACACAAATGTCATC 2551

Qy         81 LysPheThrValLysAlaSerAlaTyr 89
              ||||||||||||||||||||||||||||
Db       2552 AAATTCACTGTAAAAGCCTCCGCTTAT 2578

RESULT 2
A20606
LOCUS       A20606                  1867 bp    DNA     linear   PAT 27-JUL-1994
DEFINITION  67kD protein.
ACCESSION   A20606
VERSION     A20606.1  GI:579221
KEYWORDS    .
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 1867)
  AUTHORS   .
  TITLE     RECOMBINANT 47 AND 31kD COCOA PROTEINS AND PRECURSOR
  JOURNAL   Patent: WO 9119801-A 6 26-DEC-1991;
FEATURES             Location/Qualifiers
     source          1. .1867
                     /organism="Theobroma cacao"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3641"
     gene            14. .1714
                     /gene="67kD protein"
     CDS             14. .1714
                     /gene="67kD protein"
                     /codon_start=1
                     /protein_id="CAA01512.1"
                     /db_xref="GI:579222"
                     /translation="MVISKSPFIVLIFSLLLSFALLCSGVSAYGRKQYERDPRQQYEQ
                     CQRRCESEATEEREQEQCEQRCEREYKEQQRQQEEELQRQYQQCQGRCQEQQQGQREQ
                     QQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQRNNPYYFPKRRSFQTRFRDEE
                     GNFKILQRFAENSPPLKGINDYRLAMFEANPNTFILPHHCDAEAIYFVTNGKGTITFV
                     THENKESYNVQRGTVVSVPAGSTVYVVSQDNQEKLTIAVLALPVNSPGKYELFFPAGN
                     NKPESYYGAFSYEVLETVFNTQREKLEEILEEQRGQKRQQGQQGMFRKAKPEQIRAIS
                     QQATSPRHRGGERLAINLLSQSPVYSNQNGRFFEACPEDFSQFQNMDVAVSAFKLNQG
                     AIFVPHYNSKATFVVFVTDGYGYAQMACPHLSRQSQGSQSGRQDRREQEEESEEETFG
                     EFQQVKAPLSPGDVFVAPAGHAVTFFASKDQPLNAVAFGLNAQNNQRIFLAGKKNLVR
                     QMDSEAKELSFGVPSKLVDNIFNNPDESYFMSFSQQRQRRDERRGNPLASILDFARLF
                     "
ORIGIN

Alignment Scores:
Pred. No.:              7.33e-32        Length:         1867
Score:                  339.00          Matches:        66
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  98.51%          Mismatches:     0
Query Match:            74.02%          Indels:         0
DB:                     6               Gaps:           0
```

```
US-10-691-590-1 (1-89) x A20606 (1-1867)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1322 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 1381

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1382 GCCCCATTGTCACCTGGTGACGTCTTTGTAGCCCCGGCAGGCCATGCAGTTACATTCTTT 1441

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1442 GCATCCAAAGACCAGCCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAGAACAACCAG 1501

Qy         61 ArgIlePheLeuAlaGlyArg 67
              ||||||||||||||||||:::
Db       1502 AGAATTTTCCTTGCAGGGAAA 1522

RESULT 3
LCU551425
LOCUS       LCU551425               1248 bp    mRNA    linear   PLN 19-DEC-2003
DEFINITION  Lens culinaris vicilin partial mRNA for allergen Len c 1.0102.
ACCESSION   AJ551425
VERSION     AJ551425.1  GI:29539110
KEYWORDS    allergen Len c 1.0102.
SOURCE      Lens culinaris (lentil)
  ORGANISM  Lens culinaris
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae;
            Lens.
REFERENCE   1
  AUTHORS   Lopez-Torrejon,G., Salcedo,G., Martin-Esteban,M., Diaz-Perales,A.,
            Pascual,C.Y. and Sanchez-Monge,R.
  TITLE     Len c 1, a major allergen and vicilin from lentil seeds: Protein
            isolation and cDNA cloning
  JOURNAL   J. Allergy Clin. Immunol. 112 (6), 1208-1215 (2003)
   PUBMED   14657885
REFERENCE   2  (bases 1 to 1248)
  AUTHORS   Salcedo,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2003) Salcedo G., Biotecnologia, E.T.S.
            Agronomos, Ciudad Universitaria. Madrid, 28040, SPAIN
FEATURES             Location/Qualifiers
     source          1. .1248
                     /organism="Lens culinaris"
                     /mol_type="mRNA"
                     /cultivar="Guardena"
                     /db_xref="taxon:3864"
                     /country="Spain"
     CDS             <1. .1248
                     /function="seed storage protein"
                     /codon_start=1
                     /evidence=experimental
                     /product="allergen Len c 1.0102"
                     /protein_id="CAD87731.1"
                     /db_xref="GI:29539111"
                     /db_xref="GOA:Q84UI0"
                     /db_xref="UniProt/TrEMBL:Q84UI0"
                     /translation="SRSDQENPFIFKSNRFQTIYENENGHIRLLQKFDKRSKIFENLQ
                     NYRLLEYKSKPHTLFLPQYTDADFILVVLSGKAVLTVLNSNDRNSFNLERGDTIKLPA
                     GTIAYLANRDDNEDLRVLDLAIPVNNPGQLESFLLSGTQNQPSFLSGFNKSILEAAFN
                     TDYEEIEKVLLEDQEQEPQHRRSLRDRRQEINKENVIVKVSREQIKELSKNAKSSSKK
                     SVSSESEPFNLRSRNPIYSNKFGKFFEITPEKNPQLQDLDIFVNSVEIKEGSLLLPNY
                     NSRAIVIVTVNEGKGYFELVGQRNENQREENDDEEEQEEETSTQVQRYRAKLSPGDVF
                     VVPAGHPVAINASSDLNLIGFGINAKNNQRNFLAGEEDNVISQIQRPVKELAFPGSSR
                     EVDRLLTNQKQSHFANAQPLQIE"
     mat_peptide     1. .1245
                     /product="allergen Len c 1.0102"
                     /evidence=experimental
ORIGIN

Alignment Scores:

Pred. No.:              1.12e-13        Length:         1248
Score:                  190.00          Matches:        42
Percent Similarity:     64.10%          Conservative:   8
Best Local Similarity:  53.85%          Mismatches:     18
Query Match:            41.48%          Indels:         10
DB:                     8               Gaps:           2

US-10-691-590-1 (1-89) x LCU551425 (1-1248)

Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLysAlaPro 22
              :::  |||||||||  ||||||||||||      :::  |||:::  :::|||
Db        922 GACGATGAAGAGGAACAAGAAGAAGAGACAAGCACACAAGTGCAAAGGTATAGGGCTAAG 981

Qy         23 LeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePheAlaSer 42
              ||||||||||||||||||||||  ||||||||||||  |||         ||||||
Db        982 TTGTCTCCAGGTGATGTTTTTGTGGTTCCAGCAGGTCATCCAGTTGCCATAAATGCCTCC 1041

Qy         43 LysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGlnArgIle 62
              |||     ||||||  :::  ||||||:::||||||:::||||||||||||
Db       1042 TCAGAT------CTCAATCTAATTGGATTTGGTATCAATGCCAAAAACAATCAGAGAAAC 1095

Qy         63 PheLeuAlaGlyArgProPhePheLeuAsnHisLysGlnAsnThrAsnValIle 80
              ||||||||||||          ,     :::      |||||||||
Db       1096 TTTCTGGCAGGC----------------------GAGGAGGACAATGTCATA 1125

RESULT 4
AJ626897
LOCUS       AJ626897                1248 bp    mRNA    linear   PLN 05-FEB-2004
DEFINITION  Pisum sativum partial mRNA for Vicilin, allergen pis s 1.0101.
ACCESSION   AJ626897
VERSION     AJ626897.1  GI:42414626
KEYWORDS    allergen pis s 1.0101; Vicilin.
SOURCE      Pisum sativum (pea)
  ORGANISM  Pisum sativum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae;
            Pisum.
REFERENCE   1
  AUTHORS   Sanchez-Monge,R., Lopez-Torrejon,G., Pascual,C.Y., Varela,J.,
            Martin-Esteban,M. and Salcedo,G.
  TITLE     Vicilin and convicilin are potential major allergens from pea seeds
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1248)
  AUTHORS   Salcedo,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2004) Salcedo G., Biotecnologia microbiana.
            Bioquimica, E.T.S. Ingenieros Agronomos, Ciudad Universitaria,
            28040 Madrid, SPAIN
FEATURES             Location/Qualifiers
     source          1. .1248
                     /organism="Pisum sativum"
                     /mol_type="mRNA"
                     /cultivar="Rommel"
                     /db_xref="taxon:3888"
     CDS             <1. .1248
                     /function="seed storage protein"
                     /note="allergen pis s 1.0101"
                     /codon_start=1
                     /evidence=experimental
                     /product="Vicilin"
                     /protein_id="CAF25232.1"
                     /db_xref="GI:42414627"
                     /translation="SRSDQENPFIFKSNRFQTLYENENGHIRLLQKFDKRSKIFENLQ
                     NYRLLEYKSKPHTLFLPQYTDADFILVVLSGKATLTVLKSNDRNSFNLERGDAIKLPA
                     GTIAYLANRDDNEDLRVLDLAIPVNKPGQLQSFLLSGTQNQPSLLSGFSKNILEAAFN
                     TNYEEIEKVLLEQQEQEPQHRRSLKDRRQEINEENVIVKVSREQIEELSKNAKSSSKK
                     SVSSESGPFNLRSRNPIYSNKFGKFFEITPEKNQQLQDLDIFVNSVDIKEGSLLLPNY
                     NSRAIVIVTVTEGKGDFELVGQRNENQGKENDKEEEQEEETSKQVQLYRAKLSPGDVF
                     VIPAGHPVAINASSDLNLIGFGINAENNERNFLAGEEDNVISQVERPVKELAFPGSSH
                     EVDRLLKNQKQSYFANAQPLQRE"
ORIGIN
```

```
Alignment Scores:
Pred. No.:              4.64e-13        Length:         1248
Score:                  185.00          Matches:        41
Percent Similarity:     64.10%          Conservative:   9
Best Local Similarity:  52.56%          Mismatches:     18
Query Match:            40.39%          Indels:         10
DB:                     8               Gaps:           2

US-10-691-590-1 (1-89) x AJ626897 (1-1248)

Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLysAlaPro 22
              ::::::|||||||||   ||||||||||||      :::   |||      :::|||
Db        922 GACAAGGAAGAGGAACAAGAAGAAGAGACAAGCAAACAAGTGCAACTGTATAGAGCTAAG 981

Qy         23 LeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePheAlaSer 42
              |||||||||||||||||||||||   ||||||||||||   |||          ||||||
Db        982 TTGTCTCCAGGTGATGTTTTTGTGATTCCAGCAGGTCACCCCGTTGCCATAAATGCCTCC 1041

Qy         43 LysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGlnArgIle 62
                 |||       ||||||   :::   ||||||:::||||||:::||||||:::|||
Db       1042 TCAGAT------CTCAATCTGATTGGATTTGGTATCAATGCCGAGAACAACGAGAGAAAC 1095

Qy         63 PheLeuAlaGlyArgProPhePheLeuAsnHisLysGlnAsnThrAsnValIle 80
              ||||||||||||                          :::      |||||||||
Db       1096 TTCCTTGCAGGT----------------------GAGGAAGACAATGTCATA 1125

RESULT 5
AJ626898
LOCUS       AJ626898                1248 bp    mRNA    linear   PLN 05-FEB-2004
DEFINITION  Pisum sativum partial mRNA for Vicilin, allergen pis s 1.0102.
ACCESSION   AJ626898
VERSION     AJ626898.1  GI:42414628
KEYWORDS    allergen pis s 1.0102; Vicilin.
SOURCE      Pisum sativum (pea)
  ORGANISM  Pisum sativum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae;
            Pisum.
REFERENCE   1
  AUTHORS   Sanchez-Monge,R., Lopez-Torrejon,G., Pascual,C.Y., Varela,J.,
            Martin-Esteban,M. and Salcedo,G.
  TITLE     Vicilin and convicilin are potential major allergens from pea seeds
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1248)
  AUTHORS   Salcedo,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2004) Salcedo G., Biotecnologia microbiana.
            Bioquimica, E.T.S. Ingenieros Agronomos, Ciudad Universitaria,
            28040 Madrid, SPAIN
FEATURES             Location/Qualifiers
     source          1. .1248
                     /organism="Pisum sativum"
                     /mol_type="mRNA"
                     /cultivar="Rommel"
                     /db_xref="taxon:3888"
     CDS             <1. .1248
                     /function="seed storage protein"
                     /note="allergen pis s 1.0102"
                     /codon_start=1
                     /evidence=experimental
                     /product="Vicilin"
                     /protein_id="CAF25233.1"
                     /db_xref="GI:42414629"
                     /translation="SRSDQENPFIFKSNRFQTLYENENGHIRLLQKFDKRSKIFENLQ
                     NYRLLEYKSKPRTLFLPQCTDADFILVVLSGKATLTVLKSNDRNSFNLERGDTIKLPA
                     GTIAYLANRDDNEDLRVLDLTIPVNKPGQLQSFLLSGTQNQPSLLSGFSKNILEAAFN
                     TNYEEIEKVLLEQQEQEPQHRRSLKDRRQEINEENVIVKVSREQIEELSKNAKSSSKK
                     SVSSESGPFNLRSRNPIYSNKFGKFFEITPEKNQQLQDLDIFVNSVDIKEGSLLLPNY
                     NSRAIVIVTVTEGKGDFELVGQRNENQGKENDKEEEQEEETSKQVQLYRAKLSPGDVF
                     VIPAGHPVAINASSDLNLIGFGINAENNERNFLAGEEDNVISQVERPVKELAFPGSSH
                     EVDRLLKNQKQSYFANAQPLQRE"

ORIGIN
```

```
Alignment Scores:
Pred. No.:              4.64e-13        Length:         1248
Score:                  185.00          Matches:        41
Percent Similarity:     64.10%          Conservative:   9
Best Local Similarity:  52.56%          Mismatches:     18
Query Match:            40.39%          Indels:         10
DB:                     8               Gaps:           2

US-10-691-590-1 (1-89) x AJ626898 (1-1248)

Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLysAlaPro 22
              ::::::|||||||||   ||||||||||||      :::   |||      :::|||
Db        922 GACAAGGAAGAGGAACAAGAAGAAGAGACAAGCAAACAAGTGCAACTGTATAGAGCTAAG 981

Qy         23 LeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePheAlaSer 42
              |||||||||||||||||||||||   ||||||||||||   |||          ||||||
Db        982 TTGTCTCCAGGTGATGTTTTTGTGATTCCAGCAGGTCACCCCGTTGCCATAAATGCCTCC 1041

Qy         43 LysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGlnArgIle 62
                 |||       ||||||   :::   ||||||:::||||||:::||||||:::|||
Db       1042 TCAGAT------CTCAATCTGATTGGATTTGGTATCAATGCCGAGAACAACGAGAGAAAC 1095

Qy         63 PheLeuAlaGlyArgProPhePheLeuAsnHisLysGlnAsnThrAsnValIle 80
              ||||||||||||                          :::      |||||||||
Db       1096 TTCCTTGCAGGT----------------------GAGGAAGACAATGTCATA 1125

RESULT 6
PSVICK
LOCUS       PSVICK                  1596 bp    mRNA    linear   PLN 29-OCT-1997
DEFINITION  Pisum sativum vicK gene.
ACCESSION   X67429
VERSION     X67429.1  GI:297169
KEYWORDS    47 kDa protein; seed storage protein; vicilin; vicK gene.
SOURCE      Pisum sativum (pea)
  ORGANISM  Pisum sativum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae;
            Pisum.
REFERENCE   1  (bases 1 to 1596)
  AUTHORS   Thompson,A.J., Bown,D.P., Yaish,S. and Gatehouse,J.A.
  TITLE     Differential expression of seed storage protein genes in the Pea. . .
  JOURNAL   Biochem. Physiol. Pflanz. 187, 1-12 (1991)
REFERENCE   2  (bases 1 to 1596)
  AUTHORS   Bown,D.P.
  JOURNAL   Thesis (1992) Durham University
REFERENCE   3  (bases 1 to 1596)
  AUTHORS   Bown,D.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1992) D.P. Bown, University of Durham, Dept. of
            Biological Sciences, Science Site, South Rd, Durham, DHI 3LE, UK
FEATURES             Location/Qualifiers
     source          1. .1596
                     /organism="Pisum sativum"
                     /mol_type="mRNA"
                     /cultivar="Feltham first"
                     /db_xref="taxon:3888"
                     /clone="pLG1.63"
                     /tissue_type="cotyledon"
                     /clone_lib="plasmid pUC8"
                     /dev_stage="13-17 days post flowering"
     gene            1. .1596
                     /gene="vicK"
     CDS             13. .1329
                     /gene="vicK"
                     /codon_start=1
                     /product="vicilin 47kD protein"
                     /protein_id="CAA47814.1"
                     /db_xref="GI:297170"
                     /db_xref="GOA:Q43626"
                     /db_xref="UniProt/TrEMBL:Q43626"
                     /translation="MAATPIKPLMLLAIAFLASVCVSSRSDQENPFIFKSNRFQTLYE
```

```
                     NENGHIRLLQKFDKRSKIFENLQNYRLLEYKSKPHTLFLPQYTDADFILVVLSGKATL
                     TVLKSNDRNSFNLERGDAIKLPAGTIAYLANRDDNEDLRVLDLAIPVNKPGQLQSFLL
                     SGTQNQPSLLSGFSKNILEAAFNTNYEEIEKVLLEQQEQEPQHRRSLKDRRQEINEEN
                     VIVKVSREQIEELSKNAKSSSKKSVSSESGPFNLRSRNPIYSNKFGKFFEITPEKNQQ
                     LQDLDIFVNSVDIKEGSLLLPNYNSRAIVIVTVTEGKGDFELVGQRNENQGKENDKEE
                     EQEEETSKQVQLYRAKLSPGDVFVIPAGHPVAINASSDLNLIGFGINAENNERNFLAG
                     EEDNVISQVERPVKELAFPGSSHEVDRLLKNQKQSYFANAQPLQRE"
     polyA_site      1552
                     /gene="vicK"
ORIGIN

Alignment Scores:
Pred. No.:              6.05e-13        Length:         1596
Score:                  185.00          Matches:        41
Percent Similarity:     64.10%          Conservative:   9
Best Local Similarity:  52.56%          Mismatches:     18
Query Match:            40.39%          Indels:         10
DB:                     8               Gaps:           2

US-10-691-590-1 (1-89) x PSVICK (1-1596)

Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLysAlaPro 22
              ::::::|||||||||   ||||||||||||      :::   |||      :::|||
Db       1003 GACAAGGAAGAGGAACAAGAAGAAGAGACAAGCAAACAAGTGCAACTGTATAGAGCTAAG 1062

Qy         23 LeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePheAlaSer 42
              ||||||||||||||||||||||    ||||||||||||   |||          ||||||
Db       1063 TTGTCTCCAGGTGATGTTTTTGTGATTCCAGCAGGTCACCCCGTTGCCATAAATGCCTCC 1122

Qy         43 LysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGlnArgIle 62
                |||      ||||||   :::   ||||||:::|||||||:::|||||||:::|||
Db       1123 TCAGAT------CTCAATCTGATTGGATTTGGTATCAATGCCGAGAACAACGAGAGAAAC 1176

Qy         63 PheLeuAlaGlyArgProPhePheLeuAsnHisLysGlnAsnThrAsnValIle 80
              ||||||||||||                :::      |||||||||
Db       1177 TTCCTTGCAGGT----------------------GAGGAAGACAATGTCATA 1206

RESULT 7
PEAVIC2
LOCUS       PEAVIC2                 907 bp    mRNA    linear   PLN 27-APR-1993
DEFINITION  pea vicilin mrna, pdub2 insert.
ACCESSION   J01258
VERSION     J01258.1  GI:169165
KEYWORDS    vicilin.
SOURCE      Pisum sativum (pea)
  ORGANISM  Pisum sativum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae;
            Pisum.
REFERENCE   1  (bases 1 to 907)
  AUTHORS   Lycett,G.W., Delauney,A.J., Gatehouse,J.A., Gilroy,J., Croy,R.R.
            and Boulter,D.
  TITLE     The vicilin gene family of pea (Pisum sativum L.): a complete cDNA
            coding sequence for preprovicilin
  JOURNAL   Nucleic Acids Res. 11 (8), 2367-2380 (1983)
  MEDLINE   83220791
   PUBMED   6687941
COMMENT     Original source text: pea (pisum sativum l. var. feltham first)
            cdna to mrna.
            Submitted (26-Apr-1983) as clean copy.
FEATURES             Location/Qualifiers
     source          1. .907
                     /organism="Pisum sativum"
                     /mol_type="mRNA"
                     /db_xref="taxon:3888"
ORIGIN

Alignment Scores:
Pred. No.:              5.8e-13         Length:         907
Score:                  183.00          Matches:        41
Percent Similarity:     64.10%          Conservative:   9
Best Local Similarity:  52.56%          Mismatches:     18
Query Match:            39.96%          Indels:         10
DB:                     8               Gaps:           2

US-10-691-590-1 (1-89) x PEAVIC2 (1-907)

Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLysAlaPro 22
              |||:::||||||:::   ||||||||||||      :::   |||      ||||||
Db        452 GAGGAAGAGGAGCAAAGAGAAGAAGAGACAAAGAATCAAGTGCAAAGTTACAAAGCTAAG 511

Qy         23 LeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePheAlaSer 42
              |||:::||||||||||||||||||   ||||||||||||   |||          ||||||
Db        512 TTGACTCCAGGTGATGTTTTTGTAATTCCGGCAGGTCATCCTGTTGCCGTAAGAGCTTCA 571

Qy         43 LysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGlnArgIle 62
                :::      ||||||   :::   ||||||:::|||||||:::|||||||||||||
Db        572 TCAAAT------CTCAATTTGCTTGGATTTGGTATCAATGCTGAAAACAACCAGAGAAAC 625

Qy         63 PheLeuAlaGlyArgProPhePheLeuAsnHisLysGlnAsnThrAsnValIle 80
              ||||||||||||                :::      |||||||||
Db        626 TTCCTTGCAGGT----------------------GAGGAGGACAATGTGATA 655

RESULT 8
AY102931
LOCUS       AY102931                1732 bp    mRNA    linear   PLN 01-JUN-2003
DEFINITION  Juglans nigra vicilin seed storage protein mRNA, partial cds.
ACCESSION   AY102931
VERSION     AY102931.1  GI:31321943
KEYWORDS    .
SOURCE      Juglans nigra (black walnut)
  ORGANISM  Juglans nigra
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fagales; Juglandaceae; Juglans.
REFERENCE   1  (bases 1 to 1732)
  AUTHORS   Ling,M., Ye,J., Cockrell,G., Sampson,H., Bannon,G.A., Stanley,J.S.
            and Burks,A.W.
  TITLE     Characterization of Two Major Allergens in English and Black Walnut
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1732)
  AUTHORS   Ling,M., Ye,J., Cockrell,G., Bannon,G.A., Stanley,J.S. and
            Burks,A.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2002) Pediatrics, Arkansas Children's Hospital
            Research Institute, 1120 Marshall Street, Little Rock, AR 72202,
            USA
FEATURES             Location/Qualifiers
     source          1. .1732
                     /organism="Juglans nigra"
                     /mol_type="mRNA"
                     /db_xref="taxon:16719"
     CDS             <1. .1446
                     /note="allergen"
                     /codon_start=1
                     /product="vicilin seed storage protein"
                     /protein_id="AAM54366.1"
                     /db_xref="GI:31321944"
                     /translation="GRDRQDPQQQYHRCQRRCQIQEQSPERQRQCQQRCERQYKEQQG
                     RERGPEASPRRESKGREEEQQRHNPYYFHSQSIRSRHESEEGEVKYLERFAERTELLR
                     GIENYRVVILDANPNTFMLPHHKDAESVIVVTRGRATLTLVSQETRESFNLECGDVIR
                     VPAGATEYVINQDSNERLEMVKLLQPVNNPGQVREYYAAGAKSPDQSYLRVFSNDILV
                     AALNTPRDRLERFFDQQEQREGVIIRASQEKLRALSQHAMSAGQRPWGRRSSGGPISL
                     KSERPSYSNQFGQFFEACPEEHRQLQEMDVLVNYAEIKRGAMMVPHYNSKATVVVYVV
                     EGTGRYEMACPHVSSQSFEDQGRREQEEEESTGRFQKVTARLARGDIFVIPAGHPIAI
                     TASQNENLRLLGFGINGENNQRNFLAGQNSIINQLEREAKELSFNMPREEIEEIFESQ
                     MESYFVPTERQSRRGQGRDHPLASILGFAFF"
ORIGIN

Alignment Scores:
Pred. No.:              1.34e-12        Length:         1732
Score:                  182.50          Matches:        37
Percent Similarity:     69.44%          Conservative:   13
Best Local Similarity:  51.39%          Mismatches:     19
Query Match:            39.85%          Indels:         3
```

```
DB:                    8             Gaps:          1

US-10-691-590-1 (1-89) x AY102931 (1-1732)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||          |||||||||:::   |||   ||||||:::|||
Db       1069 AGAAGAGAACAAGAG--------GAGGAAGAGAGCACTGGGCGATTCCAGAAGGTAACG 1119

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              |||   |||:::   ||||||:::||||||   ||||||||||||   :::
Db       1120 GCTCGTTTGGCACGTGGAGATATATTTGTAATCCCGGCAGGGCATCCGATTGCGATAACT 1179

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||:::::::::   |||       :::   ||||||:::|||   :::||||||||
Db       1180 GCCTCTCAGAATGAGAATCTGCGCTTACTCGGCTTCGGCATTAATGGCGAGAACAACCAG 1239

Qy         61 ArgIlePheLeuAlaGlyArgProPhePheLeuAsn 72
              |||   ||||||||||||:::         :::|||
Db       1240 AGAAACTTCCTTGCCGGGCAAAACAGCATAATTAAC 1275

RESULT 9
AC148289
LOCUS       AC148289               114452 bp    DNA     linear   PLN 19-MAY-2004
DEFINITION  Medicago truncatula chromosome 7 BAC clone mth2-36p20, complete
            sequence.
ACCESSION   AC148289
VERSION     AC148289.2  GI:47498155
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 114452)
  AUTHORS   Town,C.D., Tallon,L.J., Arbogast,T., Althoff,R., Hine,E.,
            Monaghan,E., Smith,S.A., Utterback,T., Feldblyum,T., Koo,H. and
            Cheung,F.
  TITLE     Medicago truncatula BAC genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 114452)
  AUTHORS   Town,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2004) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 114452)
  AUTHORS   Town,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-2004) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     On May 19, 2004 this sequence version replaced gi:42627945.
FEATURES             Location/Qualifiers
     source          1. .114452
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /chromosome="7"
                     /clone="mth2-36p20"
ORIGIN

Alignment Scores:
Pred. No.:              1.89e-10        Length:         114452
Score:                  181.00          Matches:        35
Percent Similarity:     72.22%          Conservative:   17
Best Local Similarity:  48.61%          Mismatches:     18
Query Match:            39.52%          Indels:         2
DB:                     8               Gaps:           1

US-10-691-590-1 (1-89) x AC148289 (1-114452)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              :::||||||:::::::|||:::   |||||||||         :::   |||:::   :::
Db      29460 CAGAGAGAAGAAGATGAACAACAAGAAGAAGAGAGAAGCCAACAAGTGCAAAGGTACAGA 29519

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              |||   |||||||||||||||||:::|||   ||||||||||||
Db      29520 GCTAGGTTGTCTCCAGGTGATGTGTATGTAATTCCGGCAGGTCATCCAACTGTCGTTAGC 29579

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||   |||       |||:::   :::   ||||||:::||||||:::||||||:::
Db      29580 GCTTCCTCGGAT------CTCTCTTTGCTTGGATTTGGTATCAATGCTGAGAACAACGAG 29633

Qy         61 ArgIlePheLeuAlaGlyArgProPhePheLeuAsn 72
              |||   ||||||||||||:::   |||:::::::::
Db      29634 AGGAACTTCCTTGCAGGTAAACTATTTTATATACAT 29669

RESULT 10
AB019195
LOCUS       AB019195               2647 bp    mRNA    linear   PLN 18-JAN-2001
DEFINITION  Cucurbita maxima mRNA for PV100, complete cds.
ACCESSION   AB019195
VERSION     AB019195.1  GI:3808061
KEYWORDS    PV100.
SOURCE      Cucurbita maxima (winter squash)
  ORGANISM  Cucurbita maxima
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Cucurbitales; Cucurbitaceae; Cucurbita.
REFERENCE   1  (sites)
  AUTHORS   Yamada,K., Shimada,T., Kondo,M., Nishimura,M. and Hara-Nishimura,I.
  TITLE     Multiple functional proteins are produced by cleaving Asn-Gln bonds
            of a single precursor by vacuolar processing enzyme
  JOURNAL   J. Biol. Chem. 274 (4), 2563-2570 (1999)
  MEDLINE   99107919
   PUBMED   9891029
REFERENCE   2  (bases 1 to 2647)
  AUTHORS   Hara-Nishimura,I., Nishimura,M., Shimada,T. and Yamada,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-1998) Ikuko Hara-Nishimura, Kyoto University,
            Graduate School of Science, Department of Botany; Sakyo-ku, Kyoto
            606-8502, Japan (E-mail:ihnishi@gr.bot.kyoto-u.ac.jp,
            Tel:81-075-753-4125, Fax:81-075-753-4141)
FEATURES             Location/Qualifiers
     source          1. .2647
                     /organism="Cucurbita maxima"
                     /mol_type="mRNA"
                     /strain="Kurokawa Amakuri Nankin"
                     /db_xref="taxon:3661"
                     /tissue_type="cotyledon"
                     /dev_stage="middle stage of seed maturation"
     CDS             17. .2449
                     /codon_start=1
                     /product="PV100"
                     /protein_id="BAA34056.1"
                     /db_xref="GI:3808062"
                     /translation="MALSKVKLRLCLLAFTLFLACLSVGLGDKGESLSSGAGVDHDGC
                     VNRCEELKGKNVDEFAACKKGCGVNQRGSPRAEYEVCRLRCQVAERGVEQQRKCEQVC
                     EERLRERERQQRGEDVDEVERRDPENERERQERRREHERERRRRERERERERGRGRRDE
                     NERDPKREQEERQRREQERRRREQEQRERERRGERDEEDDENQRDPDWRREQERREQE
                     RRRREQEQERRERQRRGGRDDEDENQRDPDWRREQKRREQEQERRERERRGGRDDEDE
                     NQRDPDWRREQERREQERRRREREQERREREHRGGRDVEDENQRDPDWRREQERRRRE
                     EEQREREWEREHGRRGREEQRSREDERRRHERQHGGRSRVNQVAIRRTEQEQSNNPYY
                     FQEQRFQSRYRSDEGHWRVLERFSERSELLKGIKNQRLALLEARPHTFIVPHHLDAEC
                     VLLVVRGRATITTVVQEKRETRKESYNVESGDVMTIPAGTTLYLANQENEDLQIVKLV
                     QPVNNPGEFKDYLSAGGESQAYYSVFSNDVLEAALNIPRDKLERIFKQRRERGGKIVR
                     ASQEQLRALSQRATSVRKGSRGVRAPIKLESQTPVYNNQYGQMFEACPDEFPQLRRTD
                     VATSVVDIKQGGMMVPHFNSRATWVVFVSEGAGSFEMACPHIQSSQWQRGRREEERHW
                     RREEEEEREERSGRFERVAGRLSEGGVLVIPAGHPIAIMASPNENLRLVGFGINAENN
                     HRNFLAGRENIMNELDREAKELAFNVEGKQADEIFRSQRESFFTEGPEGGRRRSTERS
                     PLLSILKLAGYF"
ORIGIN

Alignment Scores:
Pred. No.:              8.78e-12        Length:         2647
Score:                  177.50          Matches:        37
Percent Similarity:     65.28%          Conservative:   10
```

```
Best Local Similarity:  51.39%          Mismatches:     24
Query Match:            38.76%          Indels:         1
DB:                     8               Gaps:           1

US-10-691-590-1 (1-89) x AB019195 (1-2647)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||:::||||||||||   ||||||   :::   |||   |||::::::|||
Db       2063 AGAAGGGAGGAAGAAGAGGAACGCGAAGAAAGAAGC---GGTAGATTCGAAAGAGTTGCC 2119

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
                 ||||||   |||   |||   |||   ||||||||||||   :::
Db       2120 GGTCGTCTATCAGAGGGCGGCGTACTCGTAATTCCGGCAGGCCATCCAATCGCCATCATG 2179

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
                 ||||||   ::::::   |||       |||   ||||||:::||||||:::||||||
Db       2180 GCTTCCCCTAATGAGAATCTCCGCTTGGTCGGGTTCGGAATCAATGCCGAAAACAACCAC 2239

Qy         61 ArgIlePheLeuAlaGlyArgProPhePheLeuAsn 72
                 |||   ||||||||||||||||         :::|||
Db       2240 AGAAACTTCCTCGCTGGGAGAGAGAACATAATGAAC 2275

RESULT 11
LCU551424
LOCUS       LCU551424               1257 bp    mRNA    linear   PLN 19-DEC-2003
DEFINITION  Lens culinaris vicilin partial mRNA for allergen Len c 1.0101.
ACCESSION   AJ551424
VERSION     AJ551424.1  GI:29539108
KEYWORDS    allergen Len c 1.0101.
SOURCE      Lens culinaris (lentil)
  ORGANISM  Lens culinaris
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae;
            Lens.
REFERENCE   1
  AUTHORS   Lopez-Torrejon,G., Salcedo,G., Martin-Esteban,M., Diaz-Perales,A.,
            Pascual,C.Y. and Sanchez-Monge,R.
  TITLE     Len c 1, a major allergen and vicilin from lentil seeds: Protein
            isolation and cDNA cloning
  JOURNAL   J. Allergy Clin. Immunol. 112 (6), 1208-1215 (2003)
   PUBMED   14657885
REFERENCE   2  (bases 1 to 1257)
  AUTHORS   Salcedo,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2003) Salcedo G., Biotecnologia, E.T.S.
            Agronomos, Ciudad Universitaria. Madrid, 28040, SPAIN
FEATURES             Location/Qualifiers
     source          1. .1257
                     /organism="Lens culinaris"
                     /mol_type="mRNA"
                     /cultivar="Guardena"
                     /db_xref="taxon:3864"
                     /country="Spain"
     CDS             <1. .1257
                     /function="seed storage protein"
                     /codon_start=1
                     /evidence=experimental
                     /product="allergen Len c 1.0101"
                     /protein_id="CAD87730.1"
                     /db_xref="GI:29539109"
                     /db_xref="GOA:Q84UI1"
                     /db_xref="UniProt/TrEMBL:Q84UI1"
                     /translation="SRSDQENPFIFKSNRFQTIYENENGHIRLLQRFDKRSKIFENLQ
                     NYRLLEYKSKPHTIFLPQFTDADFILVVLSGKAILTVLNSNDRNSFNLERGDTIKLPA
                     GTIAYLANRDDNEDLRVLDLAIPVNRPGQLQSFLLSGTQNQPSFLSGFSKNILEAAFN
                     TEYEEIEKVLLEEQEQKSQHRRSLRDKRQEITNEDVIVKVSREQIEELSKNAKSSSKK
                     SVSSESEPFNLRSRNPIYSNKFGKFFEITPEKNPQLQDLDIFVNSVEIKEGSLLLPNY
                     NSRAIVIVTVNEGKGDFELVGQRNENQQEQREENDEEEGQEEETTKQVQRYRARLSPG
                     DVLVIPAGHPVAINASSDLNLIGFGINAKNNQRNFLAGEEDNVISQIQRPVKELAFPG
                     SSREVDRLLTNQKQSHFANAQPLQIE"
     mat_peptide     1. .1254
                     /product="allergen Len c 1.0101"
                     /evidence=experimental
ORIGIN

Alignment Scores:
Pred. No.:              4.53e-12        Length:         1257
Score:                  177.00          Matches:        40
Percent Similarity:     62.82%          Conservative:   9
Best Local Similarity:  51.28%          Mismatches:     19
Query Match:            38.65%          Indels:         10
DB:                     8               Gaps:           2

US-10-691-590-1 (1-89) x LCU551424 (1-1257)

Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLysAlaPro 22
              ::::::||||||      ||||||||||||      :::   |||:::   :::|||
Db        931 GATGAGGAAGAGGGACAAGAAGAAGAGACAACCAAACAAGTGCAAAGGTATAGAGCTAGA 990

Qy         23 LeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePheAlaSer 42
              ||||||||||||||||||   |||   ||||||||||||   |||          ||||||
Db        991 TTGTCTCCAGGTGATGTTCTTGTGATTCCAGCAGGTCATCCAGTTGCCATAAATGCCTCC 1050

Qy         43 LysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGlnArgIle 62
              |||      ||||||   :::   ||||||:::||||||:::||||||||||||
Db       1051 TCAGAT------CTCAATCTAATTGGATTTGGTATCAATGCCAAAAACAATCAGAGAAAC 1104

Qy         63 PheLeuAlaGlyArgProPhePheLeuAsnHisLysGlnAsnThrAsnValIle 80
              ||||||||||||                        :::      |||||||||
Db       1105 TTTCTGGCAGGC----------------------GAGGAGGACAATGTCATA 1134

RESULT 12
LCVICNC
LOCUS       LCVICNC                 1690 bp    DNA     linear   PLN 17-FEB-1997
DEFINITION  L.culinaris gene for vicilin type C.
ACCESSION   Z48434
VERSION     Z48434.1  GI:854241
KEYWORDS    vicilin; vicilin C.
SOURCE      Lens culinaris (lentil)
  ORGANISM  Lens culinaris
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae;
            Lens.
REFERENCE   1  (bases 1 to 1690)
  AUTHORS   Saenz de Miera,L.E.
  JOURNAL   Thesis (1995) Area de Genetica, Facultad de Biologia, Universidad
            de Leon
REFERENCE   2  (bases 1 to 1690)
  AUTHORS   Saenz de Miera,L.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-1995) Saenz de Miera L.E., Universidad de Leon
            Area de Genetica, Facultad de Biologia Campus Begazana Leon Leon
            SPAIN 24071
FEATURES             Location/Qualifiers
     source          1. .1690
                     /organism="Lens culinaris"
                     /mol_type="genomic DNA"
                     /strain="ALPO"
                     /db_xref="taxon:3864"
                     /clone="pLCA902"
                     /dev_stage="seed"
     gene            join(1. .150,297. .472,631. .708,800. .1123,1376. .1688)
                     /gene="vicilin type C"
     mRNA            join(<1. .150,297. .472,631. .708,800. .1123,1376. .1688)
                     /gene="vicilin type C"
     exon            <1. .150
                     /gene="vicilin type C"
                     /number=1
     variation       49
                     /gene="vicilin type C"
                     /note="in clone pLCA228 and pLCA008 from the same plant"
                     /replace="a"
     intron          151. .296
                     /gene="vicilin type C"
```

```
                     /number=1
     exon            297. .472
                     /gene="vicilin type C"
                     /number=2
     intron          473. .630
                     /gene="vicilin type C"
                     /number=2
     exon            631. .708
                     /gene="vicilin type C"
                     /number=3
     intron          709. .799
                     /gene="vicilin type C"
                     /number=3
     variation       714
                     /gene="vicilin type C"
                     /note="in clone pLCA008 from the same plant"
                     /replace="c"
     variation       725
                     /gene="vicilin type C"
                     /note="in clone pLCA008 from the same plant"
                     /replace="c"
     variation       736
                     /gene="vicilin type C"
                     /note="in clone pLCA228 from the same plant"
                     /replace="c"
     exon            800. .1123
                     /gene="vicilin type C"
                     /number=4
     variation       822
                     /gene="vicilin type C"
                     /note="in clone pLCA228 from the same plant"
                     /replace="g"
     variation       1086
                     /gene="vicilin type C"
                     /note="in clone pLCA008 from the same plant"
                     /replace="t"
     intron          1124. .1375
                     /gene="vicilin type C"
                     /number=4
     exon            1376. .1688
                     /gene="vicilin type C"
                     /number=5
     intron          1689. .>1690
                     /number=5
ORIGIN

Alignment Scores:
Pred. No.:              1.1e-11         Length:         1690
Score:                  175.00          Matches:        37
Percent Similarity:     71.88%          Conservative:   9
Best Local Similarity:  57.81%          Mismatches:     16
Query Match:            38.21%          Indels:         2
DB:                     8               Gaps:           1

US-10-691-590-1 (1-89) x LCVICNC (1-1690)

Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLysAlaPro 22
              |||:::|||::::::   ||||||||||||      :::   |||      ||||||
Db       1505 GAGGAAGAGAAACAAAGAGAAGAAGAAACAAAGACACAAGTGCAAAGTTACAAAGCTAAG 1564

Qy         23 LeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePheAlaSer 42
              ||||||||||||||||||||||   ||||||||||||   |||         ||||||
Db       1565 TTATCTCCAGGTGATGTTTTTGTAATTCCTGCTGGTCATCCAGTTGCTGTAAAAGCTTCA 1624

Qy         43 LysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGlnArgIle 62
                 :::      ||||||   :::   ||||||:::||||||:::||||||:::|||
Db       1625 TCAAAT------CTCAATTTGCTTGGATTTGGTATCAATGCTAAGAACAACGAGAGAAAC 1678

Qy         63 PheLeuAlaGly 66
              ||||||||||||
Db       1679 TTCCTTGCAGGT 1690

RESULT 13
COTSPA
LOCUS       COTSPA                  2777 bp    DNA     linear   PLN 27-APR-1993
DEFINITION  G.hirsutum (cotton) storage protein (alpha-globulin A) gene,
            complete cds, clone gC72-a.
ACCESSION   M19378
VERSION     M19378.1  GI:167370
KEYWORDS    alpha-globulin; storage protein; vicilin.
SOURCE      Gossypium hirsutum (upland cotton)
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 2777)
  AUTHORS   Dure,L. III.
  JOURNAL   Unpublished (1988)
COMMENT     Original source text: G.hirsutum (cotton, strain Coker 201),
            cotyledon DNA, clone gC72-a.
            Draft entry and computer-readable sequence for [1] kindly provided
            by L.Dure,III., 18-MAY-1988.
FEATURES             Location/Qualifiers
     source          1. .2777
                     /organism="Gossypium hirsutum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3635"
     CDS             join(546. .1290,1380. .1555,1636. .1716,1804. .2088,
                     2182. .2712)
                     /note="vicilin precursor"
                     /codon_start=1
                     /protein_id="AAA33069.1"
                     /db_xref="GI:167371"
                     /translation="MVRNKSVFVVLLFSLFLSFGLLCSAKDFPGRRSEDDPQQRYEDC
                     RKRCQLETRGQTEQDKCEDRSETQLKEEQQRDGEDPQRRYQDCRQHCQQEERRLRPHC
                     EQSCREQYEKQQQQQPDKQFKECQQRCQWQEQRPERKQQCVKECREQYQEDPWKGERE
                     NKWREEEEEESDEGEQQQRNNPYYFHRRSFQERFREEHGNFRVLQRFADKHHLLRGIN
                     EFRIAILEANPNTFVLPHHCDAEKIYVVTNGRGTVTFVTHENKESYNVVPGVVVRIPA
                     GSTVYLANQDNREKLTIAVLHRPVNNPGQFQKFFPAGQENPQSYLRIFSREILEAVFN
                     TRSEQLDELPGGRQSHRRQQGQGMFRKASQEQIRALSQGATSPRGKGSEGYAFNLLSQ
                     TPRYSNQNGRFYEACPRNFQQQLREVDSSVVAFEINKGSIFVPHYNSKATFVVLVTEG
                     NGHVEMVCPHLSRQSSDWSSREEEEQEEQEVERRSGQYKRVRAQLSTGNLFVVPAGHP
                     VTFVASQNEDLGLLGFGLYNGQDNKRIFVAGKTNNVRQWDRQAKELAFGVESRLVDEV
                     FNNNPQESYFVSGRDRRGFDERRGSNNPLSPFLDFARLF"
     sig_peptide     546. .617
                     /note="vicilin signal peptide"
     mat_peptide     join(618. .1290,1380. .1381)
                     /product="vicilin"
     mat_peptide     join(1290,1380. .1555)
                     /product="vicilin"
     mat_peptide     1636. .1716
                     /product="vicilin"
     mat_peptide     1804. .2088
                     /product="vicilin"
     mat_peptide     2182. .2709
                     /product="vicilin"
     exon            <546. .1290
                     /note="vicilin precursor"
                     /number=1
     intron          1291. .1379
                     /note="sp intron A"
     exon            1380. .1555
                     /number=2
     intron          1556. .1635
                     /note="sp intron B"
     exon            1636. .1716
                     /number=3
     intron          1717. .1803
                     /note="sp intron C"
     exon            1804. .2088
                     /number=4
     intron          2089. .2181
                     /note="sp intron D"
     exon            2182. .>2712
                     /note="vicilin precursor"
ORIGIN      864 bp upstream of PstI site.
```

```
Alignment Scores:
Pred. No.:                1.88e-11        Length:         2777
Score:                    175.00          Matches:        35
Percent Similarity:       75.71%          Conservative:   18
Best Local Similarity:    50.00%          Mismatches:     13
Query Match:              38.21%          Indels:         4
DB:                       8               Gaps:           2

US-10-691-590-1 (1-89) x COTSPA (1-2777)

Qy          2 ArgGluGlnGluGluGluSerGluGluGluThr---------PheGlyGluPheGlnGln 18
              ||||||:::||||||:::   |||:::|||              |||:::::::::::::
Db       2305 CGAGAGGAAGAAGAACAAGAAGAACAAGAAGTTGAAAGGAGAAGTGGACAATACAAGAGA 2364

Qy         19 ValLysAlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThr 38
              |||:::|||   ||||||   |||::::::||||||   ||||||||||||   ||||||
Db       2365 GTGAGAGCCCAGTTATCAACCGGTAACCTATTTGTAGTCCCAGCAGGCCATCCAGTCACA 2424

Qy         39 PhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeu---AsnAlaGln 57
              |||   ||||||:::::::::   |||      :::   |||||||||   |||   |||
Db       2425 TTCGTTGCCTCCCAAAACGAGGACCTAGGATTGCTCGGGTTCGGACTCTACAACGGCCAA 2484

Qy         58 AsnAsnGlnArgIlePheLeuAlaGlyArg 67
              :::|||:::|||||||||:::||||||:::
Db       2485 GACAACAAGAGGATCTTCGTGGCTGGGAAA 2514

RESULT 14
SYNAGABA
LOCUS       SYNAGABA                7876 bp    DNA     linear   SYN 27-APR-1993
DEFINITION  Synthetic Gossypium hirsutum (cotton) alpha globulin A and B genes,
            complete cds.
ACCESSION   M18027
VERSION     M18027.1  GI:207904
KEYWORDS    alpha-globulin; artificial gene; vicilin.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 7876)
  AUTHORS   Chlan,C.A., Borroto,K., Kamalay,J.A. and Dure,L. III.
  TITLE     Developmental biochemistry of cottonseed embryogenesis and
            germination. XIX. Sequences and genomic organization of the alpha
            globulin (vicilin) genes of cottonseed
  JOURNAL   Plant Mol. Biol. 9, 533-546 (1987)
COMMENT     Original source text: Gossypium hirsutum tandem 1 and 3 recombinant
            DNA.
FEATURES             Location/Qualifiers
     source          1. .7876
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
     CDS             join(809. .1550,1640. .1815,1892. .1972,2056. .2337,
                     2418. .2909)
                     /note="alpha globulin B"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA72622.1"
                     /db_xref="GI:207905"
                     /translation="MVRNKSACVVLLFSLFLSFGLLCSAKDFPGRRGDDDPPKRYEDC
                     RRRCEWDTRGQKEQQQCEESCKSQYGEKDQQQRHRPEDPQRRYEECQQECRQQEERQQ
                     PQCQQRCLKRFEQEQQQSQRQFQECQQHCHQQEQRPEKKQQCVRECREKYQENPWRGE
                     REEEAEEEETEEGEQEQSHNPFHFHRRSFQSRFREEHGNFRVLQRFASRHPILRGINE
                     FRLSILEANPNTFVLPHHCDAEKIYLVTNGRGTLTFLTHENKESYNIVPGVVVKVPAG
                     STVYLANQDNKEKLIIAVLHRPVNNPGQFEEFFPAGSQRPQSYLRAFSREILEPAFNT
                     RSEQLDELFGGRQSRRRQQGQGMFRKASQEQIRALSQEATSPREKSGERFAFNLLSQT
                     PRYSNQNGRFFEACPPEFRQLRDINVTVSALQLNQGSIFVPHYNSKATFVILVTEGNG
                     YAEMVSPHLPRQSSYEEEEEEDEEEEQEQEEERRSGQYRKIRSRLSRGDIFVVPANFP
                     VTFVASQNQNLRMTGFGLYNQNINPDHNQRIFVAGKINHVRQWDSQAKELAFGVSSRL
                     VDEIFNSNPQESYFVSRQRQRASE"
     exon            <809. .1550
                     /note="alpha globulin B"
                     /number=1
     intron          1551. .1639
                     /note="alpha globulin B, intron A"
     exon            1640. .1815
                     /number=2
     intron          1816. .1891
                     /note="alpha globulin B, intron B"
     exon            1892. .1972
                     /number=3
     intron          1973. .2055
                     /note="alpha globulin B, intron C"
     exon            2056. .2337
                     /number=4
     intron          2338. .2417
                     /note="alpha globulin B, intron D"
     exon            2418. .>2909
                     /note="alpha globulin B"
                     /number=5
     CDS             join(5645. .6389,6479. .6654,6735. .6815,6903. .7187,
                     7281. .7811)
                     /note="alpha globulin A"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA72623.1"
                     /db_xref="GI:207906"
                     /translation="MVRNKSVFVVLLFSLFLSFGLLCSAKDFPGRRSEDDPQQRYEDC
                     RKRCQLETRGQTEQDKCEDRSETQLKEEQQRDGEDPQRRYQDCRQHCQQEERRLRPHC
                     EQSCREQYEKQQQQQPDKQFKECQQRCQWQEQRPERKQQCVKECREQYQEDPWKGERE
                     NKWREEEEEESDEGEQQQRNNPYYFHRRSFQERFREEHGNFRVLQRFADKHHLLRGIN
                     EFRIAILEANPNTFVLPHHCDAEKIYVVTNGRGTVTFVTHENKESYNVVPGVVVRIPA
                     GSTVYLANQDNREKLTIAVLHRPVNNPGQFQKFFPAGQENPQSYLRIFSREILEAVFN
                     TRSEQLDELPGGRQSHRRQQGQGMFRKASQEQIRALSQGATSPRGKGSEGYAFNLLSQ
                     TPRYSNQNGRFYEACPRNFQQQLREVDSSVVAFEINKGSIFVPHYNSKATFVVLVTEG
                     NGHVEMVCPHLSRQSSDWSSREEEEQEEQEVERRSGQYKRVRAQLSTGNLFVVPAGHP
                     VTFVASQNEDLGLLGFGLYNGQDNKRIFVAGKTNNVRQWDRQAKELAFGVESRLVDEV
                     FNNNPQESYFVSGRDRRGFDERRGSNNPLSPFLDFARLF"
     exon            <5645. .6389
                     /note="alpha globulin A"
                     /number=1
     intron          6390. .6478
                     /note="alpha globulin A, intron A"
     exon            6479. .6654
                     /number=2
     intron          6655. .6734
                     /note="alpha globulin A, intron B"
     exon            6735. .6815
                     /number=3
     intron          6816. .6902
                     /note="alpha globulin A, intron C"
     exon            6903. .7187
                     /number=4
     intron          7188. .7280
                     /note="alpha globulin A, intron D"
     exon            7281. .>7811
                     /note="alpha globulin A"
                     /number=5
ORIGIN

Alignment Scores:
Pred. No.:                5.78e-11        Length:         7876
Score:                    175.00          Matches:        35
Percent Similarity:       75.71%          Conservative:   18
Best Local Similarity:    50.00%          Mismatches:     13
Query Match:              38.21%          Indels:         4
DB:                       12              Gaps:           2

US-10-691-590-1 (1-89) x SYNAGABA (1-7876)

Qy          2 ArgGluGlnGluGluGluSerGluGluGluThr---------PheGlyGluPheGlnGln 18
              ||||||:::||||||:::   |||:::|||              |||:::::::::::::
Db       7404 CGAGAGGAAGAAGAACAAGAAGAACAAGAAGTTGAAAGGAGAAGTGGACAATACAAGAGA 7463

Qy         19 ValLysAlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThr 38
              |||:::|||   ||||||   |||::::::||||||   ||||||||||||   ||||||
Db       7464 GTGAGAGCCCAGTTATCAACCGGTAACCTATTTGTAGTCCCAGCAGGCCATCCAGTCACA 7523
```

```
Qy         39 PhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeu---AsnAlaGln 57
              |||   ||||||:::::::::   |||      :::   |||||||||   |||   |||
Db       7524 TTCGTTGCCTCCCAAAACGAGGACCTAGGATTGCTCGGGTTCGGACTCTACAACGGCCAA 7583

Qy         58 AsnAsnGlnArgIlePheLeuAlaGlyArg 67
              :::|||:::|||||||||:::||||||:::
Db       7584 GACAACAAGAGGATCTTCGTGGCTGGGAAA 7613

RESULT 15
AF066055
LOCUS       AF066055               2057 bp    mRNA    linear   PLN 05-JUL-2001
DEFINITION  Juglans regia vicilin-like protein precursor, mRNA, partial cds.
ACCESSION   AF066055
VERSION     AF066055.1  GI:6580761
KEYWORDS    .
SOURCE      Juglans regia (English walnut)
  ORGANISM  Juglans regia
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fagales; Juglandaceae; Juglans.
REFERENCE   1  (bases 1 to 2057)
  AUTHORS   Teuber,S.S., Jarvis,K.C., Dandekar,A.M., Peterson,W.R. and
            Ansari,A.A.
  TITLE     Identification and cloning of a complementary DNA encoding a
            vicilin-like proprotein, jug r 2, from english walnut kernel
            (Juglans regia), a major food allergen
  JOURNAL   J. Allergy Clin. Immunol. 104 (6), 1311-1320 (1999)
  MEDLINE   20057824
   PUBMED   10589017
REFERENCE   2  (bases 1 to 2057)
  AUTHORS   Teuber,S.S., Jarvis,K.C., Peterson,W.R., Dandekar,A.M. and
            Ansari,A.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-1998) Internal Medicine, University of
            California, TB 192-Division of Rheum/Allergy, Davis, CA 95616, USA
FEATURES             Location/Qualifiers
     source          1. .2057
                     /organism="Juglans regia"
                     /mol_type="mRNA"
                     /cultivar="Sunland"
                     /db_xref="taxon:51240"
                     /tissue_type="somatic embryo line"
     CDS             <1. .1784
                     /note="seed storage protein"
                     /codon_start=3
                     /product="vicilin-like protein precursor"
                     /protein_id="AAF18269.1"
                     /db_xref="GI:6580762"
                     /translation="RGRDDDDEENPRDPREQYRQCQEYCRRQGQGQRQQQQCQIRCEE
                     RLEEDQRSQEERERRRGRDVDDQNPRDPEQRYEQCQQQCERQRRGQEQTLCRRRCEQR
                     RQQEERERQRGRDRQDPQQQYHRCQRRCQIQEQSPERQRQCQQRCERQYKEQQGRERG
                     PEASPRRESRGREEEQQRHNPYYFHSQSIRSRHESEEGEVKYLERFTERTELLRGIEN
                     YRVVILDANPNTSMLPHHKDAESVAVVTRGRATLTLVSQETRESFNLECGDVIRVPAG
                     ATVYVINQDSNERLEMVKLLQPVNNPGQFREYYAAGAKSPDQSYLRVFSNDILVAALN
                     TPRDRLERFFDQQEQREGVIIRASQEKLRALSQHAMSAGQRPWGRRSSGGPISLKSES
                     PSYSNQFGQFFEACPEEHRQLQEMDVLVNYAEIKRGAMMVPHYNSKATVVVYVVEGTG
                     RYEMACPHVSSQSYEGQGRREQEEEESTGRFQKVTARLARGDIFVIPAGHPIAITASQ
                     NENLRLLGFDINGENNQRDFLAGQNNIINQLEREAKELSFNMPREEIEEIFESQMESY
                     FVPTERQSRRGQGRDHPLASILDFAFF"
ORIGIN

Alignment Scores:
Pred. No.:              1.57e-11        Length:         2057
Score:                  174.50          Matches:        36
Percent Similarity:     68.06%          Conservative:   13
Best Local Similarity:  50.00%          Mismatches:     20
Query Match:            38.10%          Indels:         3
DB:                     8               Gaps:           1

US-10-691-590-1 (1-89) x AF066055 (1-2057)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
                |||||||||||||||        |||||||||:::   |||   ||||||:::|||
Db       1407 AGAAGAGAACAAGAG---------GAGGAAGAGAGCACTGGGCGATTCCAGAAGGTAACG 1457

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              |||   |||:::   ||||||:::||||||   ||||||||||||   :::
Db       1458 GCTCGTTTGGCACGTGGAGATATATTTGTAATCCCGGCAGGGCATCCGATTGCGATAACT 1517

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||:::::::::   |||      :::   |||   :::|||   :::||||||||||
Db       1518 GCCTCTCAGAATGAGAATCTGCGCTTACTCGGCTTCGACATTAATGGCGAGAACAACCAG 1577

Qy         61 ArgIlePheLeuAlaGlyArgProPhePheLeuAsn 72
              |||   ||||||||||||:::         :::|||
Db       1578 AGAGACTTCCTTGCCGGGCAAAACAACATAATTAAC 1613

Search completed: November  5, 2005, 04:41:10
Job time : 4406.22 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 5, 2005, 00:20:21 ; Search time 568.826 Seconds
(without alignments)
926.218 Million cell updates/sec

Title: US-10-691-590-1
Perfect score: 458
Sequence: 1 RREQEEESEEETFGEFQQVK.....FLNHKQNTNVIKFTVKASAY 89

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10691590/runat_02112005_120021_13134/app_query.fasta_1.462
-DB=N_Geneseq_16Dec04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10691590_@CGN_1_1_586_@runat_02112005_120021_13134 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	339	74.0	1867	2	AAQ20377	Aaq20377 Sequence
c	2	159	34.7	562	13	ACN45981	Acn45981 Cotton pr
	3	156.5	34.2	1930	2	AAZ06382	Aaz06382 Peanut al
	4	156.5	34.2	2032	2	AAT76613	Aat76613 Peanut al
	5	156.5	34.2	2032	4	AAF90339	Aaf90339 Peanut al
	6	156.5	34.2	2032	8	ABX70603	Abx70603 Peanut Ar
	7	156.5	34.2	2032	10	ADG27462	Adg27462 Peanut Ar
	8	156.5	34.2	2041	4	AAS08537	Aas08537 DNA encod
c	9	154.5	33.7	498	13	ACN48047	Acn48047 Cotton pr
	10	151	33.0	1949	2	AAT76612	Aat76612 Peanut al
	11	151	33.0	1949	8	ABX70604	Abx70604 Peanut Ar
	12	151	33.0	1949	10	ADG27463	Adg27463 Peanut Ar
	13	149.5	32.6	1722	10	ADL18492	Adl18492 Maize glo
	14	149.5	32.6	2003	10	ADL18494	Adl18494 Maize glo
	15	143	31.2	1680	6	ABS55197	Abs55197 Glycine m
	16	142	31.0	1920	2	AAV17563	Aav17563 Coding se
	17	140.5	30.7	2171	2	AAV42311	Aav42311 Macadamia
	18	140	30.6	1251	6	ABS55196	Abs55196 Glycine m
	19	140	30.6	1254	9	ACC49562	Acc49562 Mature be
	20	140	30.6	1278	9	ACC49561	Acc49561 FLAG-tagg
	21	140	30.6	1320	2	AAV17564	Aav17564 Coding se
	22	140	30.6	1320	9	ACC49553	Acc49553 Glycine m
	23	137	29.9	1632	6	ABS55198	Abs55198 Glycine m
	24	137	29.9	1818	2	AAV17562	Aav17562 Coding se
	25	134.5	29.4	1952	4	AAS08539	Aas08539 Anaphylac
	26	134	29.3	2140	2	AAV42316	Aav42316 Macadamia
	27	133.5	29.1	1350	12	ADJ44543	Adj44543 Plant cDN
	28	131.5	28.7	2171	2	AAV42310	Aav42310 Macadamia
	29	114	24.9	1581	10	ABX11299	Abx11299 Loblolly
	30	114	24.9	1596	10	ABX16585	Abx16585 Loblolly
	31	109	23.8	591	13	ACN50453	Acn50453 Cotton ma
	32	109	23.8	659	13	ACN50255	Acn50255 Cotton no
	33	109	23.8	2332	13	ADR63468	Adr63468 Cotton cD
c	34	108	23.6	564	13	ACN49247	Acn49247 Cotton pr
	35	102	22.3	705	12	ADJ40042	Adj40042 Plant cDN
	36	100	21.8	1924	2	AAV72243	Aav72243 G. max SB
	37	99.5	21.7	1580	12	ADE80950	Ade80950 Cashew nu
	38	99.5	21.7	1730	12	ADE80949	Ade80949 Cashew nu
c	39	87	19.0	559	13	ACN48185	Acn48185 Cotton pr
	40	83.5	18.2	1600	4	AAD17529	Aad17529 Soybean G
	41	78	17.0	1500	10	ADH89318	Adh89318 A. thalia
	42	78	17.0	1500	12	ADG44114	Adg44114 A. thalia
	43	77.5	16.9	1855	4	AAF90338	Aaf90338 Peanut al
	44	76.5	16.7	338702	11	ACN44042	Acn44042 Human gen
	45	75.5	16.5	669	10	ADD16431	Add16431 DNA (SeqI

ALIGNMENTS

RESULT 1
AAQ20377
ID AAQ20377 standard; cDNA; 1867 BP.
XX
AC AAQ20377;
XX
DT 16-APR-1992 (first entry)
XX
DE Sequence of 67 kD T. cacao protein cDNA.
XX
KW Cocoa; flavour; vicilin; seed storage protein; ss.
XX
OS Theobroma cacao.
XX
FH Key Location/Qualifiers
FT CDS 14. .1714
FT /*tag= a
XX
PN WO9119801-A.
XX
PD 26-DEC-1991.
XX
PF 11-JUN-1990; 90GB-00013016.
XX
PR 11-JUN-1990; 90GB-00013016.
XX
PA (MRSC) MARS UK LTD.
XX
PI Spencer ME, Hodge R, Deakin EA, Ashton S;

```
XX
DR   WPI; 1992-024418/03.
DR   P-PSDB; AAR20181.
XX
PT   Recombinant cocoa proteins - are responsible for flavour in cocoa beans
PT   and produced in large quantities using yeast and bacterial expression
PT   vectors.
XX
PS   Claim 9; Fig 2; 59pp; English.
XX
CC   The inventors claim a 67 kD and 31 kD T. cacao protein, and fragments,
CC   and encoding DNAs. The 47 kD and 31 kD proteins are derived from the 67
CC   kD precursor. T. cacao protein cDNA was detected in a cDNA library
CC   prepared from immature cocoa beans RNA using a probe based on the AA
CC   sequence of a CNBr peptide common to the 47 kD and 31 kD polypeptides.
CC   Homology searches revealed close homologies between the 67 kD polypeptide
CC   and the vicilins, which are seed storage proteins
XX
SQ   Sequence 1867 BP; 578 A; 402 C; 460 G; 427 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              7.47e-39      Length:        1867
Score:                  339.00        Matches:       66
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  98.51%        Mismatches:    0
Query Match:            74.02%        Indels:        0
DB:                     2             Gaps:          0

US-10-691-590-1 (1-89) x AAQ20377 (1-1867)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1322 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 1381

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1382 GCCCCATTGTCACCTGGTGACGTCTTTGTAGCCCCGGCAGGCCATGCAGTTACATTCTTT 1441

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1442 GCATCCAAAGACCAGCCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAGAACAACCAG 1501

Qy         61 ArgIlePheLeuAlaGlyArg 67
              |||||||||||||||||:::
Db       1502 AGAATTTTCCTTGCAGGGAAA 1522

RESULT 2
ACN45981/c
ID   ACN45981 standard; cDNA; 562 BP.
XX
AC   ACN45981;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Cotton primed seed EST Clone ID: LIB3825-012-Q1-N6-A5, SEQ:762.
XX
KW   Cotton; plant; EST; expressed sequence tag; transgenic plant; seed;
KW   variety DP50B; library LIB3825; molecular tag; molecular marker;
KW   genetic mapping; molecular mapping; seed germination; plant growth;
KW   plant quality; plant yield; plant breeding; tissue printing; ss.
XX
OS   Gossypium hirsutum.
XX
PN   US2004123340-A1.
XX
PD   24-JUN-2004.
XX
PF   12-DEC-2001; 2001US-00021323.
XX
PR   14-DEC-2000; 2000US-0255619P.
XX
PA   (DEIK/) DEIKMAN J.
PA   (FENG/) FENG P C C.
PA   (FINC/) FINCHER K L.
PA   (ZIEG/) ZIEGLER T E.
XX
PI   Deikman J,  Feng PCC,  Fincher KL,  Ziegler TE;
XX
DR   WPI; 2004-479808/45.
XX
PT   New isolated nucleic acid molecule that encodes a plant protein or its
PT   fragment, useful for isolating a variety of agronomically significant
PT   genes associated with plant growth, quality or yield, and as molecular
PT   tags to map genes.
XX
PS   Claim 1; SEQ ID NO 762; 34pp; English.
XX
CC   The invention relates to 17880 cotton expressed sequence tags (ESTs;
CC   ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated
CC   from primed or non-primed seeds from variety DP50B, mature seeds from
CC   variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium
CC   tissue, developing fibres, carpel walls and septa from variety
CC   Nucotton33B. The invention also relates to substantially purified
CC   proteins or their fragments encoded by nucleic acid molecules of the
CC   invention, and to transformed plants having a nucleic acid construct
CC   comprising a nucleic acid of the invention. The cotton ESTs are useful as
CC   molecular tags to isolate genetic regions, to isolate genes, to map
CC   genes, to determine gene function and to determining whether genes are
CC   members of a particular gene family. The nucleic acid molecules may be
CC   used for isolating a variety of agronomically significant genes
CC   associated with plant growth, quality, yield, and could also serve as
CC   links in metabolic and catabolic pathways. The nucleic acid molecules are
CC   also useful for identifying genes important in initiating and maintaining
CC   seed germination or that may be used to mitigate stresses encountered
CC   during seed germination. The ESTs additionally enable the acquisition of
CC   promoters and cis-regulatory elements which will be useful to express
CC   agronomically significant genes in these tissues and/or other tissues,
CC   and also permits the acquisition of molecular markers useful in breeding
CC   schemes, genetic and molecular mapping, and in cloning of agronomically
CC   significant genes. The nucleic acid molecules are further useful for
CC   detecting the expression level or pattern of a protein or mRNA and for
CC   detecting the presence or quantity of a protein by tissue printing. The
CC   present sequence represents a specifically claimed EST isolated from a
CC   cotton variety DP50B primed seed cDNA library (LIB3825). The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from the US patent office at
CC   seqdata.uspto.gov/sequence.html?DocID=US20040123340
XX
SQ   Sequence 562 BP; 147 A; 137 C; 97 G; 181 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.63e-13      Length:        562
Score:                  159.00        Matches:       33
Percent Similarity:     64.47%        Conservative:  16
Best Local Similarity:  43.42%        Mismatches:    19
Query Match:            34.72%        Indels:        8
DB:                     13            Gaps:          2

US-10-691-590-1 (1-89) x ACN45981 (1-562)

Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLysAlaPro 22
              |||   :::::::::   ||||||       : |||:::::::::::::::::::::
Db        556 GAAGACCAAGACCAAGAAGAGGAAAGGAGAAGTGGCCAATACAGGAAAATAAGGTCCCGG 497

Qy         23 LeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePheAlaSer 42
              ||||||   ||||||:::|||||||   ||||||         |||||||||   ||||||
Db        496 TTGTCGCGTGGAGACATATTTGTAGTCCCAGCAAACTTTCCAGTTACATTCGTTGCATCC 437

Qy         43 LysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsn---------- 58
              ::::::|||   |||             |||||||||      ||||||
Db        436 CAAAACCAGAACCTGAGGATGACTGGGTTCGGTCTTTACAACCAGAACATTAACCCAGAC 377

Qy         59 ---AsnGlnArgIlePheLeuAlaGlyArgProPhePheLeuAsnHis 73
                 ||||||||||||||:::||||||:::         :::||||||
Db        376 CATAACCAGAGGATCTTCGTGGCTGGGAAA---------ATCAACCAT 338
```

```
RESULT 3
AAZ06382
ID   AAZ06382 standard; DNA; 1930 BP.
XX
AC   AAZ06382;
XX
DT   17-OCT-2003  (revised)
DT   09-NOV-1999  (first entry)
XX
DE   Peanut allergen, Ara h 1.
XX
KW   allergy; immune response; transgenic; allergen; epitope;
KW   immunoglobulin E; Ig E; binding site; peanut; ds.
XX
OS   Arachis hypogaea.
XX
FH   Key             Location/Qualifiers
FT   CDS             50. .1930
FT                   /*tag=  a
FT                   /product= "Ara h 1"
XX
PN   WO9938978-A1.
XX
PD   05-AUG-1999.
XX
PF   29-JAN-1999;   99WO-US002031.
XX
PR   31-JAN-1998;   98US-0073283P.
PR   13-FEB-1998;   98US-0074590P.
PR   13-FEB-1998;   98US-0074624P.
PR   13-FEB-1998;   98US-0074633P.
PR   27-AUG-1998;   98US-00141220.
XX
PA   (UYAR-) UNIV ARKANSAS.
PA   (UYNY ) UNIV NEW YORK MT SINAI SCHOOL MEDICINE.
PA   (SOSI/) SOSIN H.
XX
PI   Sosin H,  Bannon GA,  Burks AW,  Sampson HA;
XX
DR   WPI; 1999-479189/40.
DR   P-PSDB; AAY15244.
XX
PT   Modified allergen with reduced IgE binding, useful for treating e.g.
PT   allergies.
XX
PS   Disclosure; Page 32-33; 46pp; English.
XX
CC   This is the nucleotide sequence of the Ara h 1 protein from Arachis
CC   hypogea. The Ara h 1 protein has 23 IgE (Immunoglobulin E) binding
CC   epitopes, four of which are immunodominant (AAY15247, AAY15249, AAY15250
CC   and AAY15263). By modifying the IgE binding sites the ability of the
CC   allergen to provoke an immune response is downregulated. The epitopes of
CC   the IgE binding sites can therefore be modified in genetically engineered
CC   plants and animals to elicit less of an allergic response. (Updated on 17
CC   -OCT-2003 to standardise OS field)
XX
SQ   Sequence 1930 BP; 595 A; 455 C; 514 G; 366 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                3.44e-12        Length:         1930
Score:                    156.50          Matches:        34
Percent Similarity:       66.67%          Conservative:   12
Best Local Similarity:    49.28%          Mismatches:     18
Query Match:              34.17%          Indels:         5
DB:                       2               Gaps:           2

US-10-691-590-1 (1-89) x AAZ06382 (1-1930)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              |||||||||:::|||||||:::   :::|||||||      |||      :::::::|||:::
Db       1484 CGGCGGGAAGAAGAGGAGGACGAAGACGAAGAAGAGGAGGGAAGTAACAGAGAGGTGCGT 1543

Qy         21 ---------AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaVal 37
                       |||   |||      |||||||||||||:::   ||||||   |||   |||
Db       1544 AGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCTCATCCAGTA 1603

Qy         38 ThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57
                 ||||||   :::      |||:::   :::   ||||||:::|||||||:::
Db       1604 GCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATCAACGCTGAA 1657

Qy         58 AsnAsnGlnArgIlePheLeuAlaGly 66
                ||||||   ||||||||||||||||||
Db       1658 AACAACCACAGAATCTTCCTTGCAGGT 1684

RESULT 4
AAT76613
ID   AAT76613 standard; cDNA to mRNA; 2032 BP.
XX
AC   AAT76613;
XX
DT   17-OCT-2003  (revised)
DT   29-DEC-1997  (first entry)
XX
DE   Peanut allergen Ara hI cDNA clone P41b.
XX
KW   Peanut; seed storage protein; allergen; allergy; hypersensitivity;
KW   vaccine; anaphylactic shock; immunotherapy; therapy; monoclonal antibody;
KW   ELISA; analysis; Ara hI; ds.
XX
OS   Arachis hypogaea; strain Florunner.
XX
FH   Key             Location/Qualifiers
FT   CDS             50. .1930
FT                   /*tag=  a
FT   sig_peptide     50. .115
FT                   /*tag=  b
FT   mat_peptide     116. .1927
FT                   /*tag=  c
FT   polyA_signal    2005. .2010
FT                   /*tag=  d
XX
PN   WO9724139-A1.
XX
PD   10-JUL-1997.
XX
PF   23-SEP-1996;   96WO-US015222.
XX
PR   29-DEC-1995;   95US-0009455P.
PR   04-MAR-1996;   96US-00610424.
XX
PA   (UYAR-) UNIV ARKANSAS.
XX
PI   Burks AW,  Helm RM,  Cockrell G,  Stanley JS,  Bannon GA;
XX
DR   WPI; 1997-363453/33.
DR   P-PSDB; AAW22150.
XX
PT   Peanut allergens Ara hI and Ara hII - used for vaccination and in two-
PT   site monoclonal antibody based ELISA.
XX
PS   Claim 31; Page 186-188; 354pp; English.
XX
CC   This cDNA clone, designated P41b, codes for the major peanut allergen Ara
CC   hI (AAW22150), which has multiple IgE binding epitopes (see AAW24165-87).
CC   It was amplified from peanut seed cDNA using a primer (see AAT76616)
CC   based on an isolated Ara hI peptide (see AAW24206). The sequence shows
CC   significant homology with the vicilin family of seed storage proteins of
CC   other legumes. The gene is capable of producing a protein product in
CC   prokaryotic cells that is recognised by serum IgE from a large proportion
CC   of individuals with peanut hypersensitivity. Ara hI and Ara hII (see
CC   AAW24164) can be used to raise monoclonal antibodies which are used in a
CC   specific two-site MAb ELISA for the detection of Ara hI or Ara hII
CC   (claimed). IgE-binding Ara antigen epitopes may be used in vaccines to
CC   protect against allergic reactions to peanut allergens, e.g. anaphylactic
CC   shock. (Updated on 17-OCT-2003 to standardise OS field)
XX
```

```
SQ   Sequence 2032 BP; 628 A; 473 C; 530 G; 401 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              3.69e-12        Length:         2032
Score:                  156.50          Matches:        34
Percent Similarity:     66.67%          Conservative:   12
Best Local Similarity:  49.28%          Mismatches:     18
Query Match:            34.17%          Indels:         5
DB:                     2               Gaps:           2

US-10-691-590-1 (1-89) x AAT76613 (1-2032)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              |||||||||:::||||||:::   :::||||||      |||       ::::::|||:::
Db       1484 CGGCGGGAAGAAGAGGAGGACGAAGACGAAGAAGAGGAGGGAAGTAACAGAGAGGTGCGT 1543

Qy         21 ---------AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaVal 37
                      |||   |||      ||||||||||:::   ||||||   |||   |||
Db       1544 AGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCTCATCCAGTA 1603

Qy         38 ThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57
                    ||||||   :::      |||:::   :::   ||||||:::||||||:::
Db       1604 GCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATCAACGCTGAA 1657

Qy         58 AsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||   ||||||||||||||||||
Db       1658 AACAACCACAGAATCTTCCTTGCAGGT 1684

RESULT 5
AAF90339
ID   AAF90339 standard; cDNA; 2032 BP.
XX
AC   AAF90339;
XX
DT   11-SEP-2003  (revised)
DT   23-JUL-2001  (first entry)
XX
DE   Peanut allergen Ara h1 P41B cDNA.
XX
KW   Peanut; allergen; Ara h1 P41B; transgenic plant; allergy; ss.
XX
OS   Arachis hypogaea.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    676. .930
FT                   /*tag=  a
FT                   /note= "PCR amplified fragment"
XX
PN   WO200136621-A2.
XX
PD   25-MAY-2001.
XX
PF   20-NOV-2000; 2000WO-US031657.
XX
PR   19-NOV-1999;   99US-0167255P.
XX
PA   (UYAL-) UNIV ALABAMA A & M.
XX
PI   Dodo HW,  Arntzen CJ,  Konan KN,  Viquez OM;
XX
DR   WPI; 2001-355630/37.
XX
PT   Producing transgenic peanut plants that produce allergen-free seeds,
PT   useful in non-allergenic foods, by antisense or sense co-suppression of
PT   allergen-encoding genes.
XX
PS   Claim 20; Fig 5; 72pp; English.
XX
CC   The present sequence is that of peanut allergen Ara h1 P41B cDNA. A
CC   portion of this gene is homologous to the corresponding region of the
CC   peanut allergen Ara h1 P17 gene. This region has been PCR amplified,
CC   cloned in transformation vectors (pUC18 and pBI4434) in sense and
CC   antisense orientations and used to down-regulate Ara h1 P41B and Ara h1
CC   P17 allergens in peanut. This is an example of the method of the
CC   invention, which relates to the production of a peanut plant having
CC   reduced, or undetectable, allergenic protein (AP) content in its seed. A
CC   peanut plant cell is transformed with a DNA construct containing an
CC   antisense AP gene and/or sense AP gene, or their fragments, regenerated
CC   to plants, and fertile transgenic plants that produce seeds with reduced
CC   AP content are identified. The seeds are useful for preparation of
CC   allergen-free foods. (Updated on 11-SEP-2003 to standardise OS field)
XX
SQ   Sequence 2032 BP; 628 A; 473 C; 530 G; 401 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              3.69e-12        Length:         2032
Score:                  156.50          Matches:        34
Percent Similarity:     66.67%          Conservative:   12
Best Local Similarity:  49.28%          Mismatches:     18
Query Match:            34.17%          Indels:         5
DB:                     4               Gaps:           2

US-10-691-590-1 (1-89) x AAF90339 (1-2032)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              |||||||||:::||||||:::   :::||||||      |||       ::::::|||:::
Db       1484 CGGCGGGAAGAAGAGGAGGACGAAGACGAAGAAGAGGAGGGAAGTAACAGAGAGGTGCGT 1543

Qy         21 ---------AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaVal 37
                      |||   |||      ||||||||||:::   ||||||   |||   |||
Db       1544 AGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCTCATCCAGTA 1603

Qy         38 ThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57
                    ||||||   :::      |||:::   :::   ||||||:::||||||:::
Db       1604 GCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATCAACGCTGAA 1657

Qy         58 AsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||   ||||||||||||||||||
Db       1658 AACAACCACAGAATCTTCCTTGCAGGT 1684

RESULT 6
ABX70603
ID   ABX70603 standard; cDNA; 2032 BP.
XX
AC   ABX70603;
XX
DT   26-MAR-2003  (first entry)
XX
DE   Peanut Ara h1 cDNA clone P41b.
XX
KW   Peanut; ss; allergy; Ara h1; Ara h2; Ara h3; gene; IgE binding site;
KW   anaphylactic food allergen; antiallergenic; vaccine; wound healing.
XX
OS   Arachis hypogaea.
XX
PN   WO200274250-A2.
XX
PD   26-SEP-2002.
XX
PF   18-MAR-2002; 2002WO-US009108.
XX
PR   16-MAR-2001; 2001US-0276822P.
PR   18-MAR-2002; 2002US-00276822.
XX
PA   (PANA-) PANACEA PHARM.
XX
PI   Caplan M,  Sosin H,  Sampson H,  Bannon GA,  Burks WA,  Cockrell G;
PI   Compadre CM,  Connaughton C,  Helm RM,  King NE,  Kopper RA,  Maleki SJ;
PI   Rabjohn PA,  Shin DS,  Stanley JS;
XX
DR   WPI; 2003-018765/01.
DR   P-PSDB; ABU52412.
XX
PT   New modified anaphylactic food allergen, useful for preventing or
PT   treating allergic reactions associated with e.g. anaphylactic allergens.
XX
```

PS Example 3; Fig 10; 300pp; English.
XX
CC The invention relates to a modified anaphylactic food allergen has an
CC amino acid sequence that is substantially identical to that of natural
CC anaphylactic food allergen, except for a cysteine residue that has been
CC modified so that it cannot participate in the disulphide bond. The
CC modification may also comprise mutation of the IgE binding sites to
CC reduce allergenicity. Also included are: (1) a method of making a
CC modified anaphylactic food allergen; (2) a nucleotide molecule encoding
CC or for causing a site specific mutation in the modified anaphylactic food
CC allergen; (3) a transgenic plant or animal expressing the modified
CC anaphylactic food allergen; (4) a method of treating an individual by
CC reducing the clinical response to a natural anaphylactic food allergen;
CC and an isolated fragment of peanut allergen Ara h 1. The modified
CC anaphylactic food allergen is useful for preventing or treating allergic
CC reactions associated with any natural allergen such as food, insect,
CC rubber or preferably anaphylactic allergens. It is also useful for
CC treating wounds in mammals such as bovine, canine, feline, caprine,
CC ovine, porcine, murine or equine species. The present sequence is a cDNA
CC encoding a peanut allergen (e.g. Ara h1, h2 or h3)
XX
SQ Sequence 2032 BP; 628 A; 473 C; 530 G; 401 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	3.69e-12	Length:	2032
Score:	156.50	Matches:	34
Percent Similarity:	66.67%	Conservative:	12
Best Local Similarity:	49.28%	Mismatches:	18
Query Match:	34.17%	Indels:	5
DB:	8	Gaps:	2

US-10-691-590-1 (1-89) x ABX70603 (1-2032)

```
Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              |||||||||:::||||||:::   :::||||||       |||      :::::::|||::: 
Db       1484 CGGCGGGAAGAAGAGGAGGACGAAGACGAAGAAGAGGAGGGAAGTAACAGAGAGGTGCGT 1543

Qy         21 ---------AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaVal 37
                      |||   |||      ||||||||||||:::   ||||||   |||   |||  
Db       1544 AGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCTCATCCAGTA 1603

Qy         38 ThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57
                     ||||||   :::      |||:::  :::  ||||||:::|||||||::: 
Db       1604 GCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATCAACGCTGAA 1657

Qy         58 AsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||  ||||||||||||||||||
Db       1658 AACAACCACAGAATCTTCCTTGCAGGT 1684
```

RESULT 7
ADG27462
ID ADG27462 standard; cDNA; 2032 BP.
XX
AC ADG27462;
XX
DT 26-FEB-2004 (first entry)
XX
DE Peanut Ara h1 cDNA clone P41b.
XX
KW Peanut; plant; ss; gene; allergen; Ara h1; Ara h2; Ara h3;
KW glycinin A2B1a; Jug n1; antiallergic; vulnerary;
KW anaphylactic food allergen; IgE; allergy; wound.
XX
OS Arachis hypogaea.
XX
PN US2003202980-A1.
XX
PD 30-OCT-2003.
XX
PF 18-MAR-2002; 2002US-00100303.
XX
PR 29-DEC-1995; 95US-0009455P.
PR 23-SEP-1996; 96US-00717933.
PR 31-JAN-1998; 98US-0073283P.
PR 13-FEB-1998; 98US-0074590P.
PR 13-FEB-1998; 98US-0074624P.
PR 13-FEB-1998; 98US-0074633P.
PR 29-JUN-1998; 98US-00106872.
PR 27-AUG-1998; 98US-00141220.
PR 13-NOV-1998; 98US-00191593.
PR 29-JAN-1999; 99US-00240557.
PR 29-JAN-1999; 99US-00241101.
PR 11-FEB-1999; 99US-00248673.
PR 11-FEB-1999; 99US-00248674.
PR 02-MAR-1999; 99US-0122450P.
PR 02-MAR-1999; 99US-0122452P.
PR 02-MAR-1999; 99US-0122560P.
PR 02-MAR-1999; 99US-0122565P.
PR 02-MAR-1999; 99US-0122566P.
PR 11-MAR-1999; 99US-00267719.
PR 28-JAN-2000; 2000US-00494096.
PR 16-MAR-2001; 2001US-0276822P.
XX
PA (CAPL/) CAPLAN M J.
PA (SOSI/) SOSIN H B.
PA (SAMP/) SAMPSON H.
PA (BANN/) BANNON G A.
PA (BURK/) BURKS A W.
PA (COCK/) COCKRELL G.
PA (COMP/) COMPADRE C M.
PA (CONN/) CONNAUGHTON C.
PA (HELM/) HELM R M.
PA (KING/) KING N E.
PA (KOPP/) KOPPER R A.
PA (MALE/) MALEKI S J.
PA (RABJ/) RABJOHN P A.
PA (SHIN/) SHIN D S.
PA (STAN/) STANLEY J S.
XX
PI Caplan MJ, Sosin HB, Sampson H, Bannon GA, Burks AW, Cockrell G;
PI Compadre CM, Connaughton C, Helm RM, King NE, Kopper RA, Maleki SJ;
PI Rabjohn PA, Shin DS, Stanley JS;
XX
DR WPI; 2003-875632/81.
DR P-PSDB; ADG27464.
XX
PT New modified anaphylactic food allergen comprising a cysteine residue
PT which has been modified so that it cannot participate in the disulfide
PT bond, useful for treating allergic reactions or wounds.
XX
PS Example 4; SEQ ID NO 5; 194pp; English.
XX
CC The invention relates to a modified anaphylactic food allergen whose
CC amino acid sequence is substantially identical to that of a natural
CC anaphylactic food allergen. The natural anaphylactic food allergen
CC includes at least one cysteine residue that participates in a disulphide
CC bond when the natural anaphylactic food allergen is in its native
CC conformation, except that the cysteine residue has been modified so that
CC it cannot participate in the disulphide bond. Also included are a method
CC of making a modified anaphylactic food allergen, a nucleotide molecule
CC encoding a modified anaphylactic food allergen defined above, a
CC nucleotide molecule for causing a site specific mutation in a gene
CC encoding a natural anaphylactic food allergen, a transgenic plant or
CC animal expressing a modified anaphylactic food allergen defined above, a
CC method of treating an individual by reducing the clinical response to a
CC natural anaphylactic food allergen by administering a modified
CC anaphylactic food allergen and an isolated fragment of peanut allergen
CC Ara h 1, comprising at least 10 consecutive amino acids of ADG27464 or
CC ADG27465. About 10-17% of the amino acids have been modified in at least
CC one IgE epitope or all the IgE epitopes recognised when the natural
CC anaphylactic food allergen is contacted with serum IgE from individual(s)
CC allergic to the natural anaphylactic food allergen. The invention
CC discloses Peanut allergens Ara h1, Ara h2, Ara h3 (and their encoding
CC cDNAs), Soybean Glycinin A2B1a and IgE-binding epitopes of the English
CC walnut allergen Jug n1. The modified anaphylactic food allergen can be
CC used for treating allergic reactions or wounds. The present sequence

```
CC   encodes a Peanut allergen of the invention.
XX
SQ   Sequence 2032 BP; 628 A; 473 C; 530 G; 401 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              3.69e-12        Length:         2032
Score:                  156.50          Matches:        34
Percent Similarity:     66.67%          Conservative:   12
Best Local Similarity:  49.28%          Mismatches:     18
Query Match:            34.17%          Indels:         5
DB:                     10              Gaps:           2

US-10-691-590-1 (1-89) x ADG27462 (1-2032)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              |||||||||:::||||||:::   :::||||||      |||      ::::::|||:::
Db       1484 CGGCGGGAAGAAGAGGAGGACGAAGACGAAGAAGAGGAGGGAAGTAACAGAGAGGTGCGT 1543

Qy         21 ---------AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaVal 37
                       |||   |||      ||||||||||||:::   ||||||   |||   |||
Db       1544 AGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCTCATCCAGTA 1603

Qy         38 ThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57
                  ||||||   :::      |||:::   :::   ||||||:::||||||:::
Db       1604 GCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATCAACGCTGAA 1657

Qy         58 AsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||   ||||||||||||||||||
Db       1658 AACAACCACAGAATCTTCCTTGCAGGT 1684

RESULT 8
AAS08537
ID   AAS08537 standard; cDNA; 2041 BP.
XX
AC   AAS08537;
XX
DT   23-OCT-2001  (first entry)
XX
DE   DNA encoding anaphylactic antigen Ara h 1.
XX
KW   Ara h 1; anaphylactic antigen; immunoglobulin E; IgE; immunogenic;
KW   allergy; mast cell; basophil; mouse; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             50. .1930
FT                   /*tag=  a
FT                   /transl_except= (pos:902. .904, aa:Ala)
FT                   /product= "Peptide antigen Ara h 1"
XX
PN   WO200140264-A2.
XX
PD   07-JUN-2001.
XX
PF   06-DEC-2000; 2000WO-US033124.
XX
PR   06-DEC-1999;   99US-00455294.
PR   23-JUN-2000; 2000US-0213765P.
PR   27-SEP-2000; 2000US-0235797P.
XX
PA   (PANA-) PANACEA PHARM LLC.
PA   (UYAR-) UNIV ARKANSAS.
PA   (MOUN ) MOUNT SINAI SCHOOL MEDICINE NEW YORK.
XX
PI   Bannon GA,  Burks WA,  Caplan MJ,  Sampson H,  Sosin H;
XX
DR   WPI; 2001-381378/40.
DR   P-PSDB; AAU04706.
XX
PT   Antigenic fragments useful for reducing anaphylactic risk and reducing
PT   the severity and/or number of allergic symptoms in individuals sensitive
PT   to antigens, have reduced ability to bind Immunoglobulin E.
XX
PS   Disclosure; Fig 9; 100pp; English.
XX
CC   The sequence represents the coding sequence of anaphylactic antigen Ara h
CC   1. Ara h 1 is an anaphylactic antigen (A), which was used to design
CC   antigenic peptides having a reduced ability to bind immunoglobulin E
CC   (IgE) as compared with the intact (A), or having a sequence substantially
CC   identical to a portion of sequence of an antigen that includes at least
CC   one IgE binding site, where at least one IgE binding site of the peptide
CC   is altered. The antigenic peptides are used in a composition which is
CC   useful for reducing risk or severity of allergic reaction to an antigen.
CC   This is done by identifying an individual at risk of allergic reaction to
CC   an antigen by identifying prior display of allergic symptoms when exposed
CC   to the antigen, or a familial relationship with an individual who
CC   previously displayed allergic symptoms when exposed to the antigen.
CC   Following this an antigen-specific IgE present on one or more mast cells
CC   or basophils in the individual's serum is identified. The individual is
CC   then contacted with a peptide corresponding to a portion of the antigen,
CC   which is selected, formulated, and delivered so that binding of the
CC   peptide to antigen-specific IgE is reduced as compared with IgE binding
CC   of intact antigen. The composition is also useful for treating and
CC   preventing allergic reactions
XX
SQ   Sequence 2041 BP; 637 A; 473 C; 530 G; 401 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              3.72e-12        Length:         2041
Score:                  156.50          Matches:        34
Percent Similarity:     66.67%          Conservative:   12
Best Local Similarity:  49.28%          Mismatches:     18
Query Match:            34.17%          Indels:         5
DB:                     4               Gaps:           2

US-10-691-590-1 (1-89) x AAS08537 (1-2041)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              |||||||||:::||||||:::   :::||||||      |||      ::::::|||:::
Db       1484 CGGCGGGAAGAAGAGGAGGACGAAGACGAAGAAGAGGAGGGAAGTAACAGAGAGGTGCGT 1543

Qy         21 ---------AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaVal 37
                       |||   |||      ||||||||||||:::   ||||||   |||   |||
Db       1544 AGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCTCATCCAGTA 1603

Qy         38 ThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57
                  ||||||   :::      |||:::   :::   ||||||:::||||||:::
Db       1604 GCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATCAACGCTGAA 1657

Qy         58 AsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||   ||||||||||||||||||
Db       1658 AACAACCACAGAATCTTCCTTGCAGGT 1684

RESULT 9
ACN48047/c
ID   ACN48047 standard; cDNA; 498 BP.
XX
AC   ACN48047;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Cotton primed seed EST Clone ID: LIB3825-018-Q1-N6-E12, SEQ:2828.
XX
KW   Cotton; plant; EST; expressed sequence tag; transgenic plant; seed;
KW   variety DP50B; library LIB3825; molecular tag; molecular marker;
KW   genetic mapping; molecular mapping; seed germination; plant growth;
KW   plant quality; plant yield; plant breeding; tissue printing; ss.
XX
OS   Gossypium hirsutum.
XX
PN   US2004123340-A1.
XX
PD   24-JUN-2004.
XX
PF   12-DEC-2001; 2001US-00021323.
```

XX
PR 14-DEC-2000; 2000US-0255619P.
XX
PA (DEIK/) DEIKMAN J.
PA (FENG/) FENG P C C.
PA (FINC/) FINCHER K L.
PA (ZIEG/) ZIEGLER T E.
XX
PI Deikman J, Feng PCC, Fincher KL, Ziegler TE;
XX
DR WPI; 2004-479808/45.
XX
PT New isolated nucleic acid molecule that encodes a plant protein or its
PT fragment, useful for isolating a variety of agronomically significant
PT genes associated with plant growth, quality or yield, and as molecular
PT tags to map genes.
XX
PS Claim 1; SEQ ID NO 2828; 34pp; English.
XX
CC The invention relates to 17880 cotton expressed sequence tags (ESTs;
CC ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated
CC from primed or non-primed seeds from variety DP50B, mature seeds from
CC variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium
CC tissue, developing fibres, carpel walls and septa from variety
CC Nucotton33B. The invention also relates to substantially purified
CC proteins or their fragments encoded by nucleic acid molecules of the
CC invention, and to transformed plants having a nucleic acid construct
CC comprising a nucleic acid of the invention. The cotton ESTs are useful as
CC molecular tags to isolate genetic regions, to isolate genes, to map
CC genes, to determine gene function and to determining whether genes are
CC members of a particular gene family. The nucleic acid molecules may be
CC used for isolating a variety of agronomically significant genes
CC associated with plant growth, quality, yield, and could also serve as
CC links in metabolic and catabolic pathways. The nucleic acid molecules are
CC also useful for identifying genes important in initiating and maintaining
CC seed germination or that may be used to mitigate stresses encountered
CC during seed germination. The ESTs additionally enable the acquisition of
CC promoters and cis-regulatory elements which will be useful to express
CC agronomically significant genes in these tissues and/or other tissues,
CC and also permits the acquisition of molecular markers useful in breeding
CC schemes, genetic and molecular mapping, and in cloning of agronomically
CC significant genes. The nucleic acid molecules are further useful for
CC detecting the expression level or pattern of a protein or mRNA and for
CC detecting the presence or quantity of a protein by tissue printing. The
CC present sequence represents a specifically claimed EST isolated from a
CC cotton variety DP50B primed seed cDNA library (LIB3825). The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from the US patent office at
CC seqdata.uspto.gov/sequence.html?DocID=US20040123340
XX
SQ Sequence 498 BP; 156 A; 125 C; 74 G; 143 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.01e-12	Length:	498
Score:	154.50	Matches:	34
Percent Similarity:	66.20%	Conservative:	13
Best Local Similarity:	47.89%	Mismatches:	15
Query Match:	33.73%	Indels:	9
DB:	13	Gaps:	2

US-10-691-590-1 (1-89) x ACN48047 (1-498)

```
Qy         17 GlnGlnValLysAlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAla 36
              ::::::|||:::|||   |||||||||||||||:::||||||   ||||||||||||
Db        498 AAAAAAGTGAGAGCCCGGTTATCACCCGGTGACATATTTGTAGTCCCAGCAGGCCATCCA 439

Qy         37 ValThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeu---Asn 55
              |||||||||   ||||||:::::::::   |||            |||||||||   |||
Db        438 GTCACATTTGTTGCCTCCCAAAACAAAAACCTAGGATTGTTCGGGTTTGGACTTTACAAC 379

Qy         56 AlaGlnAsnAsnGlnArgIlePheLeuAlaGlyArgProPhePheLeuAsnHisLysGln 75
              |||||||||:::||||||||:::   |||:::
Db        378 GGCCAAAACAACAAAAGGATTTTTGTGGTTGGGAAA---------------------- 343
```

```
Qy         76 AsnThrAsnValIleLysPheThrValLysAla 86
              |||||||||   ::::::      :::|||
Db        342 ACCACCAATGTTAGACAATGGGTTCGCCAGGCC 310
```

RESULT 10
AAT76612
ID AAT76612 standard; cDNA to mRNA; 1949 BP.
XX
AC AAT76612;
XX
DT 17-OCT-2003 (revised)
DT 29-DEC-1997 (first entry)
XX
DE Peanut allergen Ara hI cDNA clone P17.
XX
KW Peanut; seed storage protein; allergen; allergy; hypersensitivity;
KW vaccine; anaphylactic shock; immunotherapy; therapy; monoclonal antibody;
KW ELISA; analysis; Ara hI; ds.
XX
OS Arachis hypogaea; strain Florunner.
XX

FH	Key	Location/Qualifiers
FT	CDS	3. .1847
FT		/*tag= a
FT	sig_peptide	3. .68
FT		/*tag= b
FT	mat_peptide	69. .1844
FT		/*tag= c
FT	polyA_signal	1918. .1923
FT		/*tag= d

XX
PN WO9724139-A1.
XX
PD 10-JUL-1997.
XX
PF 23-SEP-1996; 96WO-US015222.
XX
PR 29-DEC-1995; 95US-0009455P.
PR 04-MAR-1996; 96US-00610424.
XX
PA (UYAR-) UNIV ARKANSAS.
XX
PI Burks AW, Helm RM, Cockrell G, Stanley JS, Bannon GA;
XX
DR WPI; 1997-363453/33.
DR P-PSDB; AAW22149.
XX
PT Peanut allergens Ara hI and Ara hII - used for vaccination and in two-
PT site monoclonal antibody based ELISA.
XX
PS Claim 31; Page 183-185; 354pp; English.
XX
CC This cDNA clone, designated p17, codes for the major peanut allergen Ara
CC hI (AAW22149), which has multiple IgE binding epitopes (see AAW24165-87).
CC It was amplified from peanut seed cDNA using a primer (see AAT76616)
CC based on an isolated Ara hI peptide (see AAW24206). The sequence shows
CC significant homology with the vicilin family of seed storage proteins of
CC other legumes. The gene is capable of producing a protein product in
CC prokaryotic cells that is recognised by serum IgE from a large proportion
CC of individuals with peanut hypersensitivity. Ara hI and Ara hII (see
CC AAW24164) can be used to raise monoclonal antibodies which are used in a
CC specific two-site MAb ELISA for the detection of Ara hI or Ara hII
CC (claimed). IgE-binding Ara antigen epitopes may be used in vaccines to
CC protect against allergic reactions to peanut allergens, e.g. anaphylactic
CC shock. (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 1949 BP; 599 A; 455 C; 517 G; 378 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2.21e-11	Length:	1949
Score:	151.00	Matches:	36
Percent Similarity:	63.89%	Conservative:	10

```
Best Local Similarity:  50.00%        Mismatches:    18
Query Match:            32.97%        Indels:        8
DB:                     2             Gaps:          3

US-10-691-590-1 (1-89) x AAT76612 (1-1949)

Qy          1 ArgArgGluGln---------GluGluGluSerGluGluGluThrPheGlyGluPheGln 17
              ||||||||||||         |||||||||   :::||||||      |||      :::
Db       1413 CGGCGGGAACAAGAGTGGGAAGAAGAGGAGGAAGATGAAGAAGAGGAGGGAAGTAACAGA 1472

Qy         18 GlnValLys---------AlaProLeuSerProGlyAspValPheValAlaProAlaGly 34
              :::|||:::         |||   |||      ||||||||||||:::   ||||||
Db       1473 GAGGTGCGTAGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCT 1532

Qy         35 HisAlaValThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeu 54
              |||   |||         ||||||   :::      |||:::   :::   ||||||:::
Db       1533 CATCCAGTAGCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATC 1586

Qy         55 AsnAlaGlnAsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||:::||||||   ||||||||||||||||||
Db       1587 AACGCTGAAAACAACCACAGAATCTTCCTTGCAGGT 1622

RESULT 11
ABX70604
ID   ABX70604 standard; cDNA; 1949 BP.
XX
AC   ABX70604;
XX
DT   26-MAR-2003  (first entry)
XX
DE   Peanut Ara h1 cDNA clone P17.
XX
KW   Peanut; ss; allergy; Ara h1; Ara h2; Ara h3; gene; IgE binding site;
KW   anaphylactic food allergen; antiallergenic; vaccine; wound healing.
XX
OS   Arachis hypogaea.
XX
PN   WO200274250-A2.
XX
PD   26-SEP-2002.
XX
PF   18-MAR-2002; 2002WO-US009108.
XX
PR   16-MAR-2001; 2001US-0276822P.
PR   18-MAR-2002; 2002US-00276822.
XX
PA   (PANA-) PANACEA PHARM.
XX
PI   Caplan M,  Sosin H,  Sampson H,  Bannon GA,  Burks WA,  Cockrell G;
PI   Compadre CM,  Connaughton C,  Helm RM,  King NE,  Kopper RA,  Maleki SJ;
PI   Rabjohn PA,  Shin DS,  Stanley JS;
XX
DR   WPI; 2003-018765/01.
DR   P-PSDB; ABU52413.
XX
PT   New modified anaphylactic food allergen, useful for preventing or
PT   treating allergic reactions associated with e.g. anaphylactic allergens.
XX
PS   Example 3; Fig 11; 300pp; English.
XX
CC   The invention relates to a modified anaphylactic food allergen has an
CC   amino acid sequence that is substantially identical to that of natural
CC   anaphylactic food allergen, except for a cysteine residue that has been
CC   modified so that it cannot participate in the disulphide bond. The
CC   modification may also comprise mutation of the IgE binding sites to
CC   reduce allergenicity. Also included are: (1) a method of making a
CC   modified anaphylactic food allergen; (2) a nucleotide molecule encoding
CC   or for causing a site specific mutation in the modified anaphylactic food
CC   allergen; (3) a transgenic plant or animal expressing the modified
CC   anaphylactic food allergen; (4) a method of treating an individual by
CC   reducing the clinical response to a natural anaphylactic food allergen;
CC   and an isolated fragment of peanut allergen Ara h 1. The modified
CC   anaphylactic food allergen is useful for preventing or treating allergic
CC   reactions associated with any natural allergen such as food, insect,
CC   rubber or preferably anaphylactic allergens. It is also useful for
CC   treating wounds in mammals such as bovine, canine, feline, caprine,
CC   ovine, porcine, murine or equine species. The present sequence is a cDNA
CC   encoding a peanut allergen (e.g. Ara h1, h2 or h3)
XX
SQ   Sequence 1949 BP; 599 A; 455 C; 517 G; 378 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.21e-11      Length:        1949
Score:                  151.00        Matches:       36
Percent Similarity:     63.89%        Conservative:  10
Best Local Similarity:  50.00%        Mismatches:    18
Query Match:            32.97%        Indels:        8
DB:                     8             Gaps:          3

US-10-691-590-1 (1-89) x ABX70604 (1-1949)

Qy          1 ArgArgGluGln---------GluGluGluSerGluGluGluThrPheGlyGluPheGln 17
              ||||||||||||         |||||||||   :::||||||      |||      :::
Db       1413 CGGCGGGAACAAGAGTGGGAAGAAGAGGAGGAAGATGAAGAAGAGGAGGGAAGTAACAGA 1472

Qy         18 GlnValLys---------AlaProLeuSerProGlyAspValPheValAlaProAlaGly 34
              :::|||:::         |||   |||      ||||||||||||:::   ||||||
Db       1473 GAGGTGCGTAGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCT 1532

Qy         35 HisAlaValThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeu 54
              |||   |||         ||||||   :::      |||:::   :::   ||||||:::
Db       1533 CATCCAGTAGCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATC 1586

Qy         55 AsnAlaGlnAsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||:::||||||   ||||||||||||||||||
Db       1587 AACGCTGAAAACAACCACAGAATCTTCCTTGCAGGT 1622

RESULT 12
ADG27463
ID   ADG27463 standard; cDNA; 1949 BP.
XX
AC   ADG27463;
XX
DT   26-FEB-2004  (first entry)
XX
DE   Peanut Ara h1 cDNA clone P17.
XX
KW   Peanut; plant; ss; gene; allergen; Ara h1; Ara h2; Ara h3;
KW   glycinin A2B1a; Jug n1; antiallergic; vulnerary;
KW   anaphylactic food allergen; IgE; allergy; wound.
XX
OS   Arachis hypogaea.
XX
PN   US2003202980-A1.
XX
PD   30-OCT-2003.
XX
PF   18-MAR-2002; 2002US-00100303.
XX
PR   29-DEC-1995;   95US-0009455P.
PR   23-SEP-1996;   96US-00717933.
PR   31-JAN-1998;   98US-0073283P.
PR   13-FEB-1998;   98US-0074590P.
PR   13-FEB-1998;   98US-0074624P.
PR   13-FEB-1998;   98US-0074633P.
PR   29-JUN-1998;   98US-00106872.
PR   27-AUG-1998;   98US-00141220.
PR   13-NOV-1998;   98US-00191593.
PR   29-JAN-1999;   99US-00240557.
PR   29-JAN-1999;   99US-00241101.
PR   11-FEB-1999;   99US-00248673.
PR   11-FEB-1999;   99US-00248674.
PR   02-MAR-1999;   99US-0122450P.
PR   02-MAR-1999;   99US-0122452P.
PR   02-MAR-1999;   99US-0122560P.
PR   02-MAR-1999;   99US-0122565P.
```

```
PR   02-MAR-1999;   99US-0122566P.
PR   11-MAR-1999;   99US-00267719.
PR   28-JAN-2000; 2000US-00494096.
PR   16-MAR-2001; 2001US-0276822P.
XX
PA   (CAPL/) CAPLAN M J.
PA   (SOSI/) SOSIN H B.
PA   (SAMP/) SAMPSON H.
PA   (BANN/) BANNON G A.
PA   (BURK/) BURKS A W.
PA   (COCK/) COCKRELL G.
PA   (COMP/) COMPADRE C M.
PA   (CONN/) CONNAUGHTON C.
PA   (HELM/) HELM R M.
PA   (KING/) KING N E.
PA   (KOPP/) KOPPER R A.
PA   (MALE/) MALEKI S J.
PA   (RABJ/) RABJOHN P A.
PA   (SHIN/) SHIN D S.
PA   (STAN/) STANLEY J S.
XX
PI   Caplan MJ,  Sosin HB,  Sampson H,  Bannon GA,  Burks AW,  Cockrell G;
PI   Compadre CM,  Connaughton C,  Helm RM,  King NE,  Kopper RA,  Maleki SJ;
PI   Rabjohn PA,  Shin DS,  Stanley JS;
XX
DR   WPI; 2003-875632/81.
DR   P-PSDB; ADG27465.
XX
PT   New modified anaphylactic food allergen comprising a cysteine residue
PT   which has been modified so that it cannot participate in the disulfide
PT   bond, useful for treating allergic reactions or wounds.
XX
PS   Example 4; SEQ ID NO 6; 194pp; English.
XX
CC   The invention relates to a modified anaphylactic food allergen whose
CC   amino acid sequence is substantially identical to that of a natural
CC   anaphylactic food allergen. The natural anaphylactic food allergen
CC   includes at least one cysteine residue that participates in a disulphide
CC   bond when the natural anaphylactic food allergen is in its native
CC   conformation, except that the cysteine residue has been modified so that
CC   it cannot participate in the disulphide bond. Also included are a method
CC   of making a modified anaphylactic food allergen, a nucleotide molecule
CC   encoding a modified anaphylactic food allergen defined above, a
CC   nucleotide molecule for causing a site specific mutation in a gene
CC   encoding a natural anaphylactic food allergen, a transgenic plant or
CC   animal expressing a modified anaphylactic food allergen defined above, a
CC   method of treating an individual by reducing the clinical response to a
CC   natural anaphylactic food allergen by administering a modified
CC   anaphylactic food allergen and an isolated fragment of peanut allergen
CC   Ara h 1, comprising at least 10 consecutive amino acids of ADG27464 or
CC   ADG27465. About 10-17% of the amino acids have been modified in at least
CC   one IgE epitope or all the IgE epitopes recognised when the natural
CC   anaphylactic food allergen is contacted with serum IgE from individual(s)
CC   allergic to the natural anaphylactic food allergen. The invention
CC   discloses Peanut allergens Ara h1, Ara h2, Ara h3 (and their encoding
CC   cDNAs), Soybean Glycinin A2B1a and IgE-binding epitopes of the English
CC   walnut allergen Jug n1. The modified anaphylactic food allergen can be
CC   used for treating allergic reactions or wounds. The present sequence
CC   encodes a Peanut allergen of the invention.
XX
SQ   Sequence 1949 BP; 599 A; 456 C; 516 G; 378 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                2.21e-11        Length:          1949
Score:                    151.00          Matches:         36
Percent Similarity:       63.89%          Conservative:    10
Best Local Similarity:    50.00%          Mismatches:      18
Query Match:              32.97%          Indels:          8
DB:                       10              Gaps:            3

US-10-691-590-1 (1-89) x ADG27463 (1-1949)

Qy          1 ArgArgGluGln---------GluGluGluSerGluGluGluThrPheGlyGluPheGln 17
              ||||||||||||         |||||||| . :::||||||      |||      :::
Db       1413 CGGCGGGAACAAGAGTGGGAAGAAGAGGAGGAAGATGAAGAAGAGGAGGGAAGTAACAGA 1472

Qy         18 GlnValLys---------AlaProLeuSerProGlyAspValPheValAlaProAlaGly 34
              :::|||:::         |||  |||      ||||||||||||:::   ||||||
Db       1473 GAGGTGCGTAGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCT 1532

Qy         35 HisAlaValThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeu 54
              |||   |||         ||||||   :::      |||:::   :::   ||||||:::
Db       1533 CATCCAGTAGCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATC 1586

Qy         55 AsnAlaGlnAsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||:::||||||   ||||||||||||||||||
Db       1587 AACGCTGAAAACAACCACAGAATCTTCCTTGCAGGT 1622

RESULT 13
ADL18492
ID   ADL18492 standard; cDNA; 1722 BP.
XX
AC   ADL18492;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Maize globulin-1 S allele precursor encoding cDNA SEQ ID NO:1.
XX
KW   plant; plant molecular biology; commercial; research; molecular marker;
KW   plant breeding; maize; globulin-1 S allele precursor; gene; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1722
FT                   /*tag=  a
FT                   /product= "globulin-1 S allele precursor"
XX
PN   WO2003027249-A2.
XX
PD   03-APR-2003.
XX
PF   26-SEP-2002; 2002WO-US030475.
XX
PR   26-SEP-2001; 2001US-0325277P.
PR   04-APR-2002; 2002US-0370526P.
PR   04-APR-2002; 2002US-0370620P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Su W,  Andon N,  Haynes P,  Briggs SP,  Cooper B,  Glazebrook J;
PI   Goff SA,  Katagiri F,  Kreps J,  Moughamer T,  Provart N,  Ricke D;
PI   Zhu T;
XX
DR   WPI; 2003-354649/33.
DR   P-PSDB; ADL18493.
XX
PT   New isolated high protein-phenotype-associated plant nucleic acid
PT   molecule, useful in plant molecular biology, and specifically for
PT   altering protein content or level in plants, and developing molecular
PT   markers for plant breeding.
XX
PS   Claim 1; SEQ ID NO 1; 163pp; English.
XX
CC   The present invention describes an isolated nucleic acid molecule (I)
CC   comprising a plant nucleotide sequence or its complement which hybridises
CC   under low, moderate or high stringency conditions to a nucleic acid
CC   segment encoding a polypeptide comprising any one of SEQ ID NO:1-36,
CC   where the nucleotide sequence does not encode any of SEQ ID NO:1-36. The
CC   methods and compositions of the present invention are useful in plant
CC   molecular biology, particularly for commercial and research purposes, and
CC   more specifically for altering the protein content or level in plants,
CC   and to develop molecular markers for plant breeding. The present sequence
CC   encodes the maize globulin-1 S allele precursor, which is used in the
CC   exemplification of the present invention.
XX
```

```
SQ   Sequence 1722 BP; 330 A; 540 C; 645 G; 207 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              3.08e-11       Length:        1722
Score:                  149.50         Matches:       32
Percent Similarity:     63.77%         Conservative:  12
Best Local Similarity:  46.38%         Mismatches:    22
Query Match:            32.64%         Indels:        3
DB:                     10             Gaps:          2

US-10-691-590-1 (1-89) x ADL18492 (1-1722)

Qy          1 ArgArgGluGlnGluGluGluSer------GluGluGluThrPheGlyGlu---PheGln 17
              |||   |||:::|||||||||||||      |||:::|||      |||:::   :::
Db       1198 AGGAGCGAAGAAGAAGAAGAATCGTCTGAGGAGCAGGAGGAAGTCGGGCAGGGGTACCAC 1257

Qy         18 GlnValLysAlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaVal 37
                :::::|||   ||||||||||||      ||||||   ||||||||||||
Db       1258 ACCATCCGGGCGCGGCTGTCACCGGGCACGGCGTTCGTGGTGCCCGCGGGCCACCCGTTC 1317

Qy         38 ThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57
                  ||||||:::|||      |||      |||   |||   :::::::|||
Db       1318 GTCGCGGTGGCGTCCCGGGACAGCAACCTCCAGATCGTGTGCTTCGAGGTCCACGCCGAC 1377

Qy         58 AsnAsnGlnArgIlePheLeuAlaGly 66
                |||:::::::::|||||||||||
Db       1378 AGGAACGAGAAGGTGTTCCTGGCCGGC 1404

RESULT 14
ADL18494
ID   ADL18494 standard; cDNA; 2003 BP.
XX
AC   ADL18494;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Maize globulin-2 precursor encoding cDNA SEQ ID NO:3.
XX
KW   plant; plant molecular biology; commercial; research; molecular marker;
KW   plant breeding; maize; globulin-2 precursor; gene; ss.
XX
OS   Zea mays.
XX
PN   WO2003027249-A2.
XX
PD   03-APR-2003.
XX
PF   26-SEP-2002; 2002WO-US030475.
XX
PR   26-SEP-2001; 2001US-0325277P.
PR   04-APR-2002; 2002US-0370526P.
PR   04-APR-2002; 2002US-0370620P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Su W,  Andon N,  Haynes P,  Briggs SP,  Cooper B,  Glazebrook J;
PI   Goff SA,  Katagiri F,  Kreps J,  Moughamer T,  Provart N,  Ricke D;
PI   Zhu T;
XX
DR   WPI; 2003-354649/33.
DR   P-PSDB; ADL18495.
XX
PT   New isolated high protein-phenotype-associated plant nucleic acid
PT   molecule, useful in plant molecular biology, and specifically for
PT   altering protein content or level in plants, and developing molecular
PT   markers for plant breeding.
XX
PS   Claim 1; SEQ ID NO 3; 163pp; English.
XX
CC   The present invention describes an isolated nucleic acid molecule (I)
CC   comprising a plant nucleotide sequence or its complement which hybridises
CC   under low, moderate or high stringency conditions to a nucleic acid
CC   segment encoding a polypeptide comprising any one of SEQ ID NO:1-36,
CC   where the nucleotide sequence does not encode any of SEQ ID NO:1-36. The
CC   methods and compositions of the present invention are useful in plant
CC   molecular biology, particularly for commercial and research purposes, and
CC   more specifically for altering the protein content or level in plants,
CC   and to develop molecular markers for plant breeding. The present sequence
CC   encodes the maize globulin-2 precursor, which is used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 2003 BP; 403 A; 605 C; 729 G; 266 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              3.81e-11       Length:        2003
Score:                  149.50         Matches:       32
Percent Similarity:     63.77%         Conservative:  12
Best Local Similarity:  46.38%         Mismatches:    22
Query Match:            32.64%         Indels:        3
DB:                     10             Gaps:          2

US-10-691-590-1 (1-89) x ADL18494 (1-2003)

Qy          1 ArgArgGluGlnGluGluGluSer------GluGluGluThrPheGlyGlu---PheGln 17
              |||   |||:::|||||||||||||      |||:::|||      |||:::   :::
Db       1234 AGGAGCGAAGAAGAAGAAGAATCGTCTGAGGAGCAGGAGGAAGTCGGGCAGGGGTACCAC 1293

Qy         18 GlnValLysAlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaVal 37
                :::::|||   ||||||||||||      ||||||   ||||||||||||
Db       1294 ACCATCCGGGCGCGGCTGTCACCGGGCACGGCGTTCGTGGTGCCCGCGGGCCACCCGTTC 1353

Qy         38 ThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57
                  ||||||:::|||      |||      |||   |||   :::::::|||
Db       1354 GTCGCGGTGGCGTCCCGGGACAGCAACCTCCAGATCGTGTGCTTCGAGGTCCACGCCGAC 1413

Qy         58 AsnAsnGlnArgIlePheLeuAlaGly 66
                |||:::::::::|||||||||||
Db       1414 AGGAACGAGAAGGTGTTCCTGGCCGGC 1440

RESULT 15
ABS55197
ID   ABS55197 standard; DNA; 1680 BP.
XX
AC   ABS55197;
XX
DT   17-DEC-2002  (first entry)
XX
DE   Glycine max (Soybean) var. Wasesuzunari gene #3.
XX
KW   Soybean; Glycinin; atomic coordinate data; processability; soya protein;
KW   Wasesuzunari; protein co-ordinate data; gene; ds.
XX
OS   Glycine max.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1680
FT                   /*tag= a
FT                   /partial
FT                   /product= "Glycine max (Soybean) var. Wasesuzunari
FT                   protein #3"
FT                   /note= "This sequence lacks a start codon"
XX
PN   JP2002193996-A.
XX
PD   10-JUL-2002.
XX
PF   21-DEC-2000; 2000JP-00405097.
XX
PR   21-DEC-2000; 2000JP-00405097.
XX
PA   (KYOU ) UNIV KYOTO.
XX
DR   WPI; 2002-685438/74.
DR   P-PSDB; ABG71270.
XX
PT   Glycinin, beta-conglycinin and proglycinin, their crystal structures,
```

PT three dimensional coordinates, three dimensional structured and models
PT and their uses.
XX
PS Disclosure; Page 1285-1287; 1298pp; Japanese.
XX
CC The present invention relates to a new Glycinin characterised by the
CC atomic coordinate data fully defined in the specification. The structure
CC can be used for improving processability of soya protein. The present
CC nucleic acid sequence encodes the Glycine max (Soybean) var. Wasesuzunari
CC protein #3, as described in the specification
XX
SQ Sequence 1680 BP; 567 A; 385 C; 395 G; 333 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2.65e-10	Length:	1680
Score:	143.00	Matches:	31
Percent Similarity:	66.67%	Conservative:	13
Best Local Similarity:	46.97%	Mismatches:	20
Query Match:	31.22%	Indels:	2
DB:	6	Gaps:	1

US-10-691-590-1 (1-89) x ABS55197 (1-1680)

```
Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              :::   :::|||:::   :::   :::||||||         |||   ::::::   :::
Db       1294 AAAGAACAACAACAGAGGCAGCAACAGGAAGAGCAACCTTTGGAAGTGCGGAAATATAGA 1353

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              |||   ||||||      |||:::||||||   |||||||||:::   |||
Db       1354 GCTGAATTGTCTGAACAAGATATATTTGTAATCCCAGCAGGTTATCCAGTTGTGGTCAAC 1413

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              |||:::   |||      ||||||      |||||||||:::||||||:::|||||||||
Db       1414 GCTACCTCAGAT------CTGAATTTCTTTGCTTTTGGTATCAATGCCGAGAACAACCAG 1467

Qy         61 ArgIlePheLeuAlaGly 66
              |||   ||||||||||||
Db       1468 AGGAACTTCCTTGCAGGT 1485
```

Search completed: November 5, 2005, 03:06:24
Job time : 578.826 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 5, 2005, 02:34:02 ; Search time 192.704 Seconds
(without alignments)
755.710 Million cell updates/sec

Title: US-10-691-590-1
Perfect score: 458
Sequence: 1 RREQEEESEEETFGEFQQVK.....FLNHKQNTNVIKFTVKASAY 89

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10691590/runat_02112005_120023_13203/app_query.fasta_1.462
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10691590_@CGN_1_1_93_@runat_02112005_120023_13203 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	339	74.0	1867	1	US-07-955-905A-1	Sequence 1, Appli
2	156.5	34.2	1930	4	US-09-106-872A-3	Sequence 3, Appli
3	156.5	34.2	2032	4	US-09-106-872A-21	Sequence 21, Appl
4	156.5	34.2	2032	4	US-09-191-593-5	Sequence 5, Appli
5	156.5	34.2	2041	4	US-09-106-872A-23	Sequence 23, Appl
6	151	33.0	1949	4	US-09-106-872A-19	Sequence 19, Appl
7	151	33.0	1949	4	US-09-191-593-4	Sequence 4, Appli
8	142	31.0	1920	3	US-09-108-010B-2	Sequence 2, Appli
9	142	31.0	1920	4	US-09-758-652-2	Sequence 2, Appli
10	142	31.0	1920	4	US-10-684-651-2	Sequence 2, Appli
11	140	30.6	1320	3	US-09-108-010B-3	Sequence 3, Appli
12	140	30.6	1320	4	US-09-758-652-3	Sequence 3, Appli
13	140	30.6	1320	4	US-10-684-651-3	Sequence 3, Appli
14	137	29.9	1818	3	US-09-108-010B-1	Sequence 1, Appli
15	137	29.9	1818	4	US-09-758-652-1	Sequence 1, Appli
16	137	29.9	1818	4	US-10-684-651-1	Sequence 1, Appli
17	119	26.0	1590	3	US-09-323-195A-7	Sequence 7, Appli
18	114	24.9	1581	3	US-09-323-195A-16	Sequence 16, Appl
19	114	24.9	1596	3	US-09-323-195A-19	Sequence 19, Appl
20	109.5	23.9	1413	3	US-09-323-195A-8	Sequence 8, Appli
21	100	21.8	1924	3	US-09-424-283-5	Sequence 5, Appli
22	71	15.5	116966	4	US-09-949-016-17557	Sequence 17557, A
23	69.5	15.2	2677	4	US-09-949-016-1924	Sequence 1924, Ap
24	69.5	15.2	19237	4	US-09-949-016-13666	Sequence 13666, A
25	68.5	15.0	1182	1	US-07-640-476-4	Sequence 4, Appli
26	68.5	15.0	1185	6	5290690-1	Patent No. 5290690
27	68.5	15.0	1185	6	5290690-1	Patent No. 5290690
28	68.5	15.0	1566	6	5290690-3	Patent No. 5290690
29	68.5	15.0	1566	6	5290690-3	Patent No. 5290690
30	67.5	14.7	805	4	US-09-270-767-1535	Sequence 1535, Ap
31	67.5	14.7	805	4	US-09-270-767-16817	Sequence 16817, A
32	67	14.6	1551	3	US-09-108-010B-15	Sequence 15, Appl
33	67	14.6	1551	4	US-09-758-652-15	Sequence 15, Appl
34	67	14.6	1551	4	US-10-684-651-15	Sequence 15, Appl
35	67	14.6	5066	4	US-09-824-574-1	Sequence 1, Appli
36	67	14.6	1664976	4	US-08-916-421B-1	Sequence 1, Appli
37	67	14.6	1664976	4	US-09-692-570-1	Sequence 1, Appli
38	65	14.2	1785	4	US-09-614-221A-386	Sequence 386, App
39	65	14.2	2103	4	US-08-933-711B-1	Sequence 1, Appli
40	64.5	14.1	411	4	US-09-107-433-881	Sequence 881, App
c 41	64.5	14.1	690	4	US-09-252-991A-7236	Sequence 7236, Ap
42	64.5	14.1	1300	3	US-08-971-782-3	Sequence 3, Appli
43	64.5	14.1	1300	3	US-09-309-026-3	Sequence 3, Appli
44	64.5	14.1	1380	4	US-09-583-110-2548	Sequence 2548, Ap
45	64.5	14.1	1638	4	US-09-902-540-8719	Sequence 8719, Ap

ALIGNMENTS

RESULT 1
US-07-955-905A-1
; Sequence 1, Application US/07955905A
; Patent No. 5770433
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: RECOMBINANT 47 AND 31 kD COCOA PROTEINS AND
; TITLE OF INVENTION: PRECURSOR
; NUMBER OF SEQUENCES: 28
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/955,905A
; FILING DATE: 21-JAN-1993
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1867 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Theobroma cacao
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14..1711
US-07-955-905A-1

Alignment Scores:
Pred. No.: 5.71e-40 Length: 1867
Score: 339.00 Matches: 66
Percent Similarity: 100.00% Conservative: 1

```
Best Local Similarity:  98.51%          Mismatches:     0
Query Match:            74.02%          Indels:         0
DB:                     1               Gaps:           0

US-10-691-590-1 (1-89) x US-07-955-905A-1 (1-1867)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1322 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 1381

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1382 GCCCCATTGTCACCTGGTGACGTCTTTGTAGCCCCGGCAGGCCATGCAGTTACATTCTTT 1441

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1442 GCATCCAAAGACCAGCCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAGAACAACCAG 1501

Qy         61 ArgIlePheLeuAlaGlyArg 67
              |||||||||||||||||:::
Db       1502 AGAATTTTCCTTGCAGGGAAA 1522

RESULT 2
US-09-106-872A-3
; Sequence 3, Application US/09106872A
; Patent No. 6486311
; GENERAL INFORMATION:
;  APPLICANT:  Burks Jr., A. Wesley
;  APPLICANT:  Stanley, J. Steven
;  APPLICANT:  Cockrell, Gael
;  APPLICANT:  King, Nina E.
;  APPLICANT:  Sampson, Hugh A.
;  APPLICANT:  Helm, Ricki M.
;  APPLICANT:  Bannon, Gary A.
;  TITLE OF INVENTION:  Peanut Allergens and Methods
;  FILE REFERENCE:  HS 103 CIP
;  CURRENT APPLICATION NUMBER: US/09/106,872A
;  CURRENT FILING DATE:  1999-06-29
;  PRIOR APPLICATION NUMBER:  PCT/US96/15222
;  PRIOR FILING DATE:  1996-09-23
;  NUMBER OF SEQ ID NOS:  23
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 1930
;   TYPE: DNA
;   ORGANISM: Arachis hypogaea
US-09-106-872A-3

Alignment Scores:
Pred. No.:              5.91e-13        Length:         1930
Score:                  156.50          Matches:        34
Percent Similarity:     66.67%          Conservative:   12
Best Local Similarity:  49.28%          Mismatches:     18
Query Match:            34.17%          Indels:         5
DB:                     4               Gaps:           2

US-10-691-590-1 (1-89) x US-09-106-872A-3 (1-1930)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              |||||||||:::||||||:::   :::|||||||      |||      ::::::|||:::
Db       1484 CGGCGGGAAGAAGAGGAGGACGAAGACGAAGAAGAGGAGGGAAGTAACAGAGAGGTGCGT 1543

Qy         21 ---------AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaVal 37
                      |||   |||      ||||||||||||:::   ||||||   |||   |||
Db       1544 AGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCTCATCCAGTA 1603

Qy         38 ThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57
                    ||||||   :::      |||:::   :::   ||||||:::||||||:::
Db       1604 GCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATCAACGCTGAA 1657

Qy         58 AsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||   ||||||||||||||||||
Db       1658 AACAACCACAGAATCTTCCTTGCAGGT 1684

RESULT 3
US-09-106-872A-21
; Sequence 21, Application US/09106872A
; Patent No. 6486311
; GENERAL INFORMATION:
;  APPLICANT:  Burks Jr., A. Wesley
;  APPLICANT:  Stanley, J. Steven
;  APPLICANT:  Cockrell, Gael
;  APPLICANT:  King, Nina E.
;  APPLICANT:  Sampson, Hugh A.
;  APPLICANT:  Helm, Ricki M.
;  APPLICANT:  Bannon, Gary A.
;  TITLE OF INVENTION:  Peanut Allergens and Methods
;  FILE REFERENCE:  HS 103 CIP
;  CURRENT APPLICATION NUMBER: US/09/106,872A
;  CURRENT FILING DATE:  1999-06-29
;  PRIOR APPLICATION NUMBER:  PCT/US96/15222
;  PRIOR FILING DATE:  1996-09-23
;  NUMBER OF SEQ ID NOS:  23
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 21
;   LENGTH: 2032
;   TYPE: DNA
;   ORGANISM: Arachis hypogaea
US-09-106-872A-21

Alignment Scores:
Pred. No.:              6.37e-13        Length:         2032
Score:                  156.50          Matches:        34
Percent Similarity:     66.67%          Conservative:   12
Best Local Similarity:  49.28%          Mismatches:     18
Query Match:            34.17%          Indels:         5
DB:                     4               Gaps:           2

US-10-691-590-1 (1-89) x US-09-106-872A-21 (1-2032)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              |||||||||:::||||||:::   :::||||||      |||      ::::::|||:::
Db       1484 CGGCGGGAAGAAGAGGAGGACGAAGACGAAGAAGAGGAGGGAAGTAACAGAGAGGTGCGT 1543

Qy         21 ---------AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaVal 37
                      |||   |||      ||||||||||||:::   ||||||   |||   |||
Db       1544 AGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCTCATCCAGTA 1603

Qy         38 ThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57
                    ||||||   :::      |||:::   :::   ||||||:::||||||:::
Db       1604 GCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATCAACGCTGAA 1657

Qy         58 AsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||   ||||||||||||||||||
Db       1658 AACAACCACAGAATCTTCCTTGCAGGT 1684

RESULT 4
US-09-191-593-5
; Sequence 5, Application US/09191593
; Patent No. 6835824
;  GENERAL INFORMATION:
;    APPLICANT:  BURKS, A Wesley, HELM, Ricki M,
;    APPLICANT:  COCKRELL, Gael, STANLEY, J Steven,
;    APPLICANT:  BANNON, Gary A
;    TITLE OF INVENTION:  PEANUT ALLERGENS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  67
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Head, Johnson & Kachigian
;      STREET:  112 W. Center St., Suite 230
;      CITY:  Fayetteville
;      STATE:  Arkansas  AR
;      COUNTRY:  United States of America
;      ZIP:  72701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb
```

; MEDIUM TYPE: storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS 6.2
; SOFTWARE: Wordperfect 6.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/191,593
; FILING DATE: 13 NOVEMBER 1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/717,933
; FILING DATE: 23 SEPTEMBER 1996
; APPLICATION NUMBER: US 07/998,377
; FILING DATE: 30 DECEMBER 1992
; APPLICATION NUMBER: US 08/158,704
; FILING DATE: 29 NOVEMBER 1993
; APPLICATION NUMBER: US 60/009,455
; FILING DATE: 29 DECEMBER 1995
; APPLICATION NUMBER: US 08/610,424
; FILING DATE: 04 MARCH 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ALEXANDER, DANIEL R
; REGISTRATION NUMBER: 32,604
; REFERENCE/DOCKET NUMBER: ARK00895601B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (501) 582-9111
; TELEFAX: (501) 521-4931
; TELEX: No. 6835824 applicable
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2032 bases
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; DESCRIPTION: identified as Ara h I Beta P41b
; HYPOTHETICAL: No
; ANTI-SENSE: No
; FRAGMENT TYPE: No. 6835824 applicable
; ORIGINAL SOURCE:
; ORGANISM: Arachis hypogaea
; STRAIN: Florunner
; INDIVIDUAL ISOLATE: Clone P41b
; DEVELOPMENTAL STAGE: Seed
; HAPLOTYPE: No. 6835824 applicable
; TISSUE TYPE: Seed mRNA, cDNA library
; CELL TYPE: No. 6835824 applicable
; CELL LINE: No. 6835824 applicable
; ORGANELLE: No. 6835824 applicable
; IMMEDIATE SOURCE:
; LIBRARY: Florunner seed cDNA expression
; LIBRARY: library in Uni-ZAP XR vector
; CLONE: P41b
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: No. 6835824 applicable
; MAP POSITION: No. 6835824 applicable
; UNITS: No. 6835824 applicable
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 50..1930
; IDENTIFICATION METHOD: By agreement with
; IDENTIFICATION METHOD: protein information and established
; IDENTIFICATION METHOD: consensus sequence
; OTHER INFORMATION: Seed storage protein and
; OTHER INFORMATION: allergen
; PUBLICATION INFORMATION:
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; ISSUE:
; PAGES:
; DATE:
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:

US-09-191-593-5

Alignment Scores:

Pred. No.:	6.37e-13	Length:	2032
Score:	156.50	Matches:	34
Percent Similarity:	66.67%	Conservative:	12
Best Local Similarity:	49.28%	Mismatches:	18
Query Match:	34.17%	Indels:	5
DB:	4	Gaps:	2

US-10-691-590-1 (1-89) x US-09-191-593-5 (1-2032)

```
Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||:::||||||:::   :::||||||      |||      ::::::|||:::
Db       1484 CGGCGGGAAGAAGAGGAGGACGAAGACGAAGAAGAGGAGGGAAGTAACAGAGAGGTGCGT 1543

Qy         21 ---------AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaVal 37
                      |||   |||      ||||||||||||:::   ||||||   |||   |||
Db       1544 AGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCTCATCCAGTA 1603

Qy         38 ThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57
                    ||||||   :::      |||:::   :::   ||||||:::||||||:::
Db       1604 GCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATCAACGCTGAA 1657

Qy         58 AsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||   ||||||||||||||||||
Db       1658 AACAACCACAGAATCTTCCTTGCAGGT 1684
```

RESULT 5
US-09-106-872A-23
; Sequence 23, Application US/09106872A
; Patent No. 6486311
; GENERAL INFORMATION:
; APPLICANT: Burks Jr., A. Wesley
; APPLICANT: Stanley, J. Steven
; APPLICANT: Cockrell, Gael
; APPLICANT: King, Nina E.
; APPLICANT: Sampson, Hugh A.
; APPLICANT: Helm, Ricki M.
; APPLICANT: Bannon, Gary A.
; TITLE OF INVENTION: Peanut Allergens and Methods
; FILE REFERENCE: HS 103 CIP
; CURRENT APPLICATION NUMBER: US/09/106,872A
; CURRENT FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: PCT/US96/15222
; PRIOR FILING DATE: 1996-09-23
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 2041
; TYPE: DNA
; ORGANISM: Arachis hypogaea
US-09-106-872A-23

Alignment Scores:

Pred. No.:	6.41e-13	Length:	2041
Score:	156.50	Matches:	34
Percent Similarity:	66.67%	Conservative:	12
Best Local Similarity:	49.28%	Mismatches:	18
Query Match:	34.17%	Indels:	5
DB:	4	Gaps:	2

US-10-691-590-1 (1-89) x US-09-106-872A-23 (1-2041)

```
Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||:::||||||:::   :::||||||      |||      ::::::|||:::
Db       1484 CGGCGGGAAGAAGAGGAGGACGAAGACGAAGAAGAGGAGGGAAGTAACAGAGAGGTGCGT 1543

Qy         21 ---------AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaVal 37
                      |||   |||      ||||||||||||:::   ||||||   |||   |||
```

```
Db       1544 AGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCTCATCCAGTA 1603

Qy         38 ThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57
                  ||||||   :::      |||:::   :::   ||||||:::||||||:::
Db       1604 GCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATCAACGCTGAA 1657

Qy         58 AsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||   ||||||||||||||||||
Db       1658 AACAACCACAGAATCTTCCTTGCAGGT 1684

RESULT 6
US-09-106-872A-19
; Sequence 19, Application US/09106872A
; Patent No. 6486311
; GENERAL INFORMATION:
;  APPLICANT:  Burks Jr., A. Wesley
;  APPLICANT:  Stanley, J. Steven
;  APPLICANT:  Cockrell, Gael
;  APPLICANT:  King, Nina E.
;  APPLICANT:  Sampson, Hugh A.
;  APPLICANT:  Helm, Ricki M.
;  APPLICANT:  Bannon, Gary A.
;  TITLE OF INVENTION:  Peanut Allergens and Methods
;  FILE REFERENCE:  HS 103 CIP
;  CURRENT APPLICATION NUMBER: US/09/106,872A
;  CURRENT FILING DATE:  1999-06-29
;  PRIOR APPLICATION NUMBER:  PCT/US96/15222
;  PRIOR FILING DATE:  1996-09-23
;  NUMBER OF SEQ ID NOS:  23
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 19
;   LENGTH: 1949
;   TYPE: DNA
;   ORGANISM: Arachis hypogaea
US-09-106-872A-19

Alignment Scores:
Pred. No.:              3.9e-12       Length:          1949
Score:                  151.00        Matches:         36
Percent Similarity:     63.89%        Conservative:    10
Best Local Similarity:  50.00%        Mismatches:      18
Query Match:            32.97%        Indels:          8
DB:                     4             Gaps:            3

US-10-691-590-1 (1-89) x US-09-106-872A-19 (1-1949)

Qy          1 ArgArgGluGln---------GluGluGluSerGluGluGluThrPheGlyGluPheGln 17
              ||||||||||||         |||||||||   :::||||||      |||      :::
Db       1413 CGGCGGGAACAAGAGTGGGAAGAAGAGGAGGAAGATGAAGAAGAGGAGGGAAGTAACAGA 1472

Qy         18 GlnValLys---------AlaProLeuSerProGlyAspValPheValAlaProAlaGly 34
              :::|||:::         |||   |||       ||||||||||||:::   ||||||
Db       1473 GAGGTGCGTAGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCT 1532

Qy         35 HisAlaValThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeu 54
              |||   |||         ||||||   :::       |||:::   :::   ||||||:::
Db       1533 CATCCAGTAGCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATC 1586

Qy         55 AsnAlaGlnAsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||:::||||||   ||||||||||||||||||
Db       1587 AACGCTGAAAACAACCACAGAATCTTCCTTGCAGGT 1622

RESULT 7
US-09-191-593-4
; Sequence 4, Application US/09191593
; Patent No. 6835824
;  GENERAL INFORMATION:
;    APPLICANT:  BURKS, A Wesley, HELM, Ricki M,
;    APPLICANT:  COCKRELL, Gael, STANLEY, J Steven,
;    APPLICANT:  BANNON, Gary A
;    TITLE OF INVENTION:  PEANUT ALLERGENS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  67
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Head, Johnson & Kachigian
;      STREET:  112 W. Center St., Suite 230
;      CITY:  Fayetteville
;      STATE:  Arkansas  AR
;      COUNTRY:  United States of America
;      ZIP:  72701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  MS-DOS 6.2
;      SOFTWARE:  Wordperfect 6.0C
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/191,593
;      FILING DATE:  13 NOVEMBER 1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/717,933
;      FILING DATE:  23 SEPTEMBER 1996
;      APPLICATION NUMBER:  US 07/998,377
;      FILING DATE:  30 DECEMBER 1992
;      APPLICATION NUMBER:  US 08/158,704
;      FILING DATE:  29 NOVEMBER 1993
;      APPLICATION NUMBER:  US 60/009,455
;      FILING DATE:  29 DECEMBER 1995
;      APPLICATION NUMBER:  US 08/610,424
;      FILING DATE:  04 MARCH 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ALEXANDER, DANIEL R
;      REGISTRATION NUMBER:  32,604
;      REFERENCE/DOCKET NUMBER:  ARK00895601B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (501)  582-9111
;      TELEFAX:  (501) 521-4931
;      TELEX:  No. 6835824 applicable
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1949 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;      DESCRIPTION:  identified as Ara h I Alpha P17
;    HYPOTHETICAL:  No
;    ANTI-SENSE:  No
;    FRAGMENT TYPE:  No. 6835824 applicable
;    ORIGINAL SOURCE:
;      ORGANISM:  Arachis hypogaea
;      STRAIN:  Florunner
;      INDIVIDUAL ISOLATE:  Clone P17
;      DEVELOPMENTAL STAGE:  Seed
;      HAPLOTYPE:  No. 6835824 applicable
;      TISSUE TYPE:  Seed mRNA, cDNA library
;      CELL TYPE:  No. 6835824 applicable
;      CELL LINE:  No. 6835824 applicable
;      ORGANELLE:  No. 6835824 applicable
;    IMMEDIATE SOURCE:
;      LIBRARY:  Florunner seed cDNA expression
;      LIBRARY:  library in Uni-ZAP  XR vector
;      CLONE:  P17
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:  No. 6835824 applicable
;      MAP POSITION:  No. 6835824 applicable
;      UNITS:  No. 6835824 applicable
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..1847
;      IDENTIFICATION METHOD:  By agreement with
;      IDENTIFICATION METHOD:  protein information and established
;      IDENTIFICATION METHOD:  consensus sequence
;      OTHER INFORMATION:  Seed storage protein and
;      OTHER INFORMATION:  allergen
```

```
;    PUBLICATION INFORMATION:
;      AUTHORS:
;      TITLE:
;      JOURNAL:
;      VOLUME:
;      ISSUE:
;      PAGES:
;      DATE:
;      DOCUMENT NUMBER:
;      FILING DATE:
;      PUBLICATION DATE:
;      RELEVANT RESIDUES IN SEQ ID NO:
US-09-191-593-4

Alignment Scores:
Pred. No.:                3.9e-12        Length:          1949
Score:                    151.00         Matches:         36
Percent Similarity:       63.89%         Conservative:    10
Best Local Similarity:    50.00%         Mismatches:      18
Query Match:              32.97%         Indels:          8
DB:                       4              Gaps:            3

US-10-691-590-1 (1-89) x US-09-191-593-4 (1-1949)

Qy          1 ArgArgGluGln---------GluGluGluSerGluGluGluThrPheGlyGluPheGln 17
              ||||||||||||         |||||||||   :::||||||      |||      :::
Db       1413 CGGCGGGAACAAGAGTGGGAAGAAGAGGAGGAAGATGAAGAAGAGGAGGGAAGTAACAGA 1472

Qy         18 GlnValLys---------AlaProLeuSerProGlyAspValPheValAlaProAlaGly 34
              :::|||:::         |||   |||      ||||||||||:::  ||||||
Db       1473 GAGGTGCGTAGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCT 1532

Qy         35 HisAlaValThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeu 54
              |||   |||         ||||||   :::      |||:::   :::   ||||||:::
Db       1533 CATCCAGTAGCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATC 1586

Qy         55 AsnAlaGlnAsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||:::||||||   ||||||||||||||||||
Db       1587 AACGCTGAAAACAACCACAGAATCTTCCTTGCAGGT 1622

RESULT 8
US-09-108-010B-2
; Sequence 2, Application US/09108010B
; Patent No. 6362399
;   GENERAL INFORMATION:
;        APPLICANT: ANTHONY JOHN KINNEY
;                   GARY MICHAEL FADER
;        TITLE OF INVENTION: SUPPRESSION OF SPECIFIC
;                            CLASSES OF SOYBEAN SEED
;                            PROTEIN GENES
;        NUMBER OF SEQUENCES: 22
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
;             STREET: 1007 MARKET STREET
;             CITY: WILMINGTON
;             STATE: DELAWARE
;             COUNTRY: UNITED STATES OF AMERICA
;             ZIP: 19898
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: DISKETTE, 3.50 INCH
;             COMPUTER: IBM PC COMPATIBLE
;             OPERATING SYSTEM: MICROSOFT WINDOWS 95
;             SOFTWARE: MICROFSOFT WORD 7.0A
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/108,010B
;             FILING DATE: 30-Jun-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/019,940
;             FILING DATE: JUNE 14,1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: LYNNE M. CHRISTENBURY
;             REGISTRATION NUMBER: 30,971
;             REFERENCE/DOCKET NUMBER: BB-1071-A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 302-992-5481
;             TELEFAX: 302-773-0164
;   INFORMATION FOR SEQ ID NO: 2:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1920 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-108-010B-2

Alignment Scores:
Pred. No.:                8.19e-11       Length:          1920
Score:                    142.00         Matches:         31
Percent Similarity:       66.67%         Conservative:    13
Best Local Similarity:    46.97%         Mismatches:      20
Query Match:              31.00%         Indels:          2
DB:                       3              Gaps:            1

US-10-691-590-1 (1-89) x US-09-108-010B-2 (1-1920)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              :::   :::|||:::   :::   :::||||||      |||   ::::::   :::
Db       1534 AAAGAACAACAACAGAGGCAGCAACAGGAAGAGCAACCTTTGGAAGTGCGGAAATATAGA 1593

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              |||   ||||||      ||:::||||||   |||||||||:::   |||
Db       1594 GCTGAATTGTCTGAACAAGATATATTTGTAATCCCAGCAGGTTATCCAGTTATGGTCAAC 1653

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              |||:::   |||      ||||||      |||||||||:::||||||:::|||||||||
Db       1654 GCTACCTCAGAT------CTGAATTTCTTTGCTTTTGGTATCAATGCCGAGAACAACCAG 1707

Qy         61 ArgIlePheLeuAlaGly 66
              |||   ||||||||||||
Db       1708 AGGAACTTCCTTGCAGGT 1725

RESULT 9
US-09-758-652-2
; Sequence 2, Application US/09758652
; Patent No. 6703544
;   GENERAL INFORMATION:
;        APPLICANT: ANTHONY JOHN KINNEY
;                   GARY MICHAEL FADER
;        TITLE OF INVENTION: SUPPRESSION OF SPECIFIC
;                            CLASSES OF SOYBEAN SEED
;                            PROTEIN GENES
;        NUMBER OF SEQUENCES: 21
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
;             STREET: 1007 MARKET STREET
;             CITY: WILMINGTON
;             STATE: DELAWARE
;             COUNTRY: UNITED STATES OF AMERICA
;             ZIP: 19898
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: DISKETTE, 3.50 INCH
;             COMPUTER: IBM PC COMPATIBLE
;             OPERATING SYSTEM: MICROSOFT WINDOWS 95
;             SOFTWARE: MICROFSOFT WORD 7.0A
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/758,652
;             FILING DATE: 11-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/019,940
;             FILING DATE: JUNE 14,1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: LYNNE M. CHRISTENBURY
;             REGISTRATION NUMBER: 30,971
```

```
;            REFERENCE/DOCKET NUMBER: BB-1071-A
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 302-992-5481
;            TELEFAX: 302-773-0164
;   INFORMATION FOR SEQ ID NO: 2:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1920 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-758-652-2

Alignment Scores:
Pred. No.:              8.19e-11      Length:        1920
Score:                  142.00        Matches:       31
Percent Similarity:     66.67%        Conservative:  13
Best Local Similarity:  46.97%        Mismatches:    20
Query Match:            31.00%        Indels:        2
DB:                     4             Gaps:          1

US-10-691-590-1 (1-89) x US-09-758-652-2 (1-1920)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              :::   :::|||:::   :::   :::||||||          |||   ::::::   :::
Db       1534 AAAGAACAACAACAGAGGCAGCAACAGGAAGAGCAACCTTTGGAAGTGCGGAAATATAGA 1593

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              |||   ||||||      |||:::||||||   |||||||||:::   |||
Db       1594 GCTGAATTGTCTGAACAAGATATATTTGTAATCCCAGCAGGTTATCCAGTTATGGTCAAC 1653

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              |||:::   |||      ||||||      |||||||||:::||||||:::|||||||||
Db       1654 GCTACCTCAGAT------CTGAATTTCTTTGCTTTTGGTATCAATGCCGAGAACAACCAG 1707

Qy         61 ArgIlePheLeuAlaGly 66
              |||   ||||||||||||
Db       1708 AGGAACTTCCTTGCAGGT 1725

RESULT 10
US-10-684-651-2
; Sequence 2, Application US/10684651
; Patent No. 6828491
;   GENERAL INFORMATION:
;       APPLICANT: ANTHONY JOHN KINNEY
;                  GARY MICHAEL FADER
;       TITLE OF INVENTION: SUPPRESSION OF SPECIFIC
;                           CLASSES OF SOYBEAN SEED
;                           PROTEIN GENES
;       NUMBER OF SEQUENCES: 21
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
;            STREET: 1007 MARKET STREET
;            CITY: WILMINGTON
;            STATE: DELAWARE
;            COUNTRY: UNITED STATES OF AMERICA
;            ZIP: 19898
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: DISKETTE, 3.50 INCH
;            COMPUTER: IBM PC COMPATIBLE
;            OPERATING SYSTEM: MICROSOFT WINDOWS 95
;            SOFTWARE: MICROFSOFT WORD 7.0A
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/684,651
;            FILING DATE: 14-Oct-2003
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/09/758,652
;            FILING DATE: 11-Jan-2001
;            APPLICATION NUMBER: 60/019,940
;            FILING DATE: JUNE 14,1996
;       ATTORNEY/AGENT INFORMATION:
;            NAME: LYNNE M. CHRISTENBURY
;            REGISTRATION NUMBER: 30,971
;            REFERENCE/DOCKET NUMBER: BB-1071-A
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 302-992-5481
;            TELEFAX: 302-773-0164
;   INFORMATION FOR SEQ ID NO: 2:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1920 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-684-651-2

Alignment Scores:
Pred. No.:              8.19e-11      Length:        1920
Score:                  142.00        Matches:       31
Percent Similarity:     66.67%        Conservative:  13
Best Local Similarity:  46.97%        Mismatches:    20
Query Match:            31.00%        Indels:        2
DB:                     4             Gaps:          1

US-10-691-590-1 (1-89) x US-10-684-651-2 (1-1920)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              :::   :::|||:::   :::   :::||||||          |||   ::::::   :::
Db       1534 AAAGAACAACAACAGAGGCAGCAACAGGAAGAGCAACCTTTGGAAGTGCGGAAATATAGA 1593

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              |||   ||||||      |||:::||||||   |||||||||:::   |||
Db       1594 GCTGAATTGTCTGAACAAGATATATTTGTAATCCCAGCAGGTTATCCAGTTATGGTCAAC 1653

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              |||:::   |||      ||||||      |||||||||:::||||||:::|||||||||
Db       1654 GCTACCTCAGAT------CTGAATTTCTTTGCTTTTGGTATCAATGCCGAGAACAACCAG 1707

Qy         61 ArgIlePheLeuAlaGly 66
              |||   ||||||||||||
Db       1708 AGGAACTTCCTTGCAGGT 1725

RESULT 11
US-09-108-010B-3
; Sequence 3, Application US/09108010B
; Patent No. 6362399
;   GENERAL INFORMATION:
;       APPLICANT: ANTHONY JOHN KINNEY
;                  GARY MICHAEL FADER
;       TITLE OF INVENTION: SUPPRESSION OF SPECIFIC
;                           CLASSES OF SOYBEAN SEED
;                           PROTEIN GENES
;       NUMBER OF SEQUENCES: 22
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
;            STREET: 1007 MARKET STREET
;            CITY: WILMINGTON
;            STATE: DELAWARE
;            COUNTRY: UNITED STATES OF AMERICA
;            ZIP: 19898
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: DISKETTE, 3.50 INCH
;            COMPUTER: IBM PC COMPATIBLE
;            OPERATING SYSTEM: MICROSOFT WINDOWS 95
;            SOFTWARE: MICROFSOFT WORD 7.0A
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/108,010B
;            FILING DATE: 30-Jun-1998
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/019,940
;            FILING DATE: JUNE 14,1996
;       ATTORNEY/AGENT INFORMATION:
```

```
;            NAME: LYNNE M. CHRISTENBURY
;            REGISTRATION NUMBER: 30,971
;            REFERENCE/DOCKET NUMBER: BB-1071-A
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 302-992-5481
;            TELEFAX: 302-773-0164
;   INFORMATION FOR SEQ ID NO: 3:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1320 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-108-010B-3

Alignment Scores:
Pred. No.:              9.44e-11      Length:         1320
Score:                  140.00        Matches:        33
Percent Similarity:     59.26%        Conservative:   15
Best Local Similarity:  40.74%        Mismatches:     23
Query Match:            30.57%        Indels:         10
DB:                     3             Gaps:           2

US-10-691-590-1 (1-89) x US-09-108-010B-3 (1-1320)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              :::   :::|||:::::::::   :::||||||        |||   |||:::   :::
Db        934 AAAGAACAACAACAGAAGCAGAAACAGGAAGAGGAACCTTTGGAAGTGCAAAGGTACAGA 993

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              |||   ||||||      ||||||||||   ||||||      |||       |||
Db        994 GCTGAATTGTCTGAAGACGATGTATTTGTAATTCCAGCA------GCTTATCCATTTGTC 1047

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
                :::               ||||||   :::|||||||||:::||||||:::|||||||||
Db       1048 GTCAACGCTACCTCAAACCTCAATTTCCTTGCTTTTGGTATCAATGCTGAGAACAACCAG 1107

Qy         61 ArgIlePheLeuAlaGlyArgProPhePheLeuAsnHisLysGlnAsnThrAsnValIle 80
              |||   ||||||||||||                        :::      ||||||:::
Db       1108 AGGAACTTCCTTGCAGGC----------------------GAGAAAGACAATGTGGTA 1143

Qy         81 Lys 81
              :::
Db       1144 AGG 1146

RESULT 12
US-09-758-652-3
; Sequence 3, Application US/09758652
; Patent No. 6703544
;   GENERAL INFORMATION:
;       APPLICANT: ANTHONY JOHN KINNEY
;                  GARY MICHAEL FADER
;       TITLE OF INVENTION: SUPPRESSION OF SPECIFIC
;                           CLASSES OF SOYBEAN SEED
;                           PROTEIN GENES
;       NUMBER OF SEQUENCES: 21
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
;            STREET: 1007 MARKET STREET
;            CITY: WILMINGTON
;            STATE: DELAWARE
;            COUNTRY: UNITED STATES OF AMERICA
;            ZIP: 19898
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: DISKETTE, 3.50 INCH
;            COMPUTER: IBM PC COMPATIBLE
;            OPERATING SYSTEM: MICROSOFT WINDOWS 95
;            SOFTWARE: MICROFSOFT WORD 7.0A
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/758,652
;            FILING DATE: 11-Jan-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/019,940
;            FILING DATE: JUNE 14,1996
;       ATTORNEY/AGENT INFORMATION:
;            NAME: LYNNE M. CHRISTENBURY
;            REGISTRATION NUMBER: 30,971
;            REFERENCE/DOCKET NUMBER: BB-1071-A
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 302-992-5481
;            TELEFAX: 302-773-0164
;   INFORMATION FOR SEQ ID NO: 3:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1320 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-758-652-3

Alignment Scores:
Pred. No.:              9.44e-11      Length:         1320
Score:                  140.00        Matches:        33
Percent Similarity:     59.26%        Conservative:   15
Best Local Similarity:  40.74%        Mismatches:     23
Query Match:            30.57%        Indels:         10
DB:                     4             Gaps:           2

US-10-691-590-1 (1-89) x US-09-758-652-3 (1-1320)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              :::   :::|||:::::::::   :::||||||        |||   |||:::   :::
Db        934 AAAGAACAACAACAGAAGCAGAAACAGGAAGAGGAACCTTTGGAAGTGCAAAGGTACAGA 993

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              |||   ||||||      ||||||||||   ||||||      |||       |||
Db        994 GCTGAATTGTCTGAAGACGATGTATTTGTAATTCCAGCA------GCTTATCCATTTGTC 1047

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
                :::               ||||||   :::|||||||||:::||||||:::|||||||||
Db       1048 GTCAACGCTACCTCAAACCTCAATTTCCTTGCTTTTGGTATCAATGCTGAGAACAACCAG 1107

Qy         61 ArgIlePheLeuAlaGlyArgProPhePheLeuAsnHisLysGlnAsnThrAsnValIle 80
              |||   ||||||||||||                        :::      ||||||:::
Db       1108 AGGAACTTCCTTGCAGGC----------------------GAGAAAGACAATGTGGTA 1143

Qy         81 Lys 81
              :::
Db       1144 AGG 1146

RESULT 13
US-10-684-651-3
; Sequence 3, Application US/10684651
; Patent No. 6828491
;   GENERAL INFORMATION:
;       APPLICANT: ANTHONY JOHN KINNEY
;                  GARY MICHAEL FADER
;       TITLE OF INVENTION: SUPPRESSION OF SPECIFIC
;                           CLASSES OF SOYBEAN SEED
;                           PROTEIN GENES
;       NUMBER OF SEQUENCES: 21
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
;            STREET: 1007 MARKET STREET
;            CITY: WILMINGTON
;            STATE: DELAWARE
;            COUNTRY: UNITED STATES OF AMERICA
;            ZIP: 19898
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: DISKETTE, 3.50 INCH
;            COMPUTER: IBM PC COMPATIBLE
;            OPERATING SYSTEM: MICROSOFT WINDOWS 95
;            SOFTWARE: MICROFSOFT WORD 7.0A
```

```
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/10/684,651
;              FILING DATE: 14-Oct-2003
;              CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: US/09/758,652
;              FILING DATE: 11-Jan-2001
;              APPLICATION NUMBER: 60/019,940
;              FILING DATE: JUNE 14,1996
;         ATTORNEY/AGENT INFORMATION:
;              NAME: LYNNE M. CHRISTENBURY
;              REGISTRATION NUMBER: 30,971
;              REFERENCE/DOCKET NUMBER: BB-1071-A
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: 302-992-5481
;              TELEFAX: 302-773-0164
;    INFORMATION FOR SEQ ID NO: 3:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 1320 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: cDNA
;         SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-684-651-3

Alignment Scores:
Pred. No.:              9.44e-11        Length:         1320
Score:                  140.00          Matches:        33
Percent Similarity:     59.26%          Conservative:   15
Best Local Similarity:  40.74%          Mismatches:     23
Query Match:            30.57%          Indels:         10
DB:                     4               Gaps:           2

US-10-691-590-1 (1-89) x US-10-684-651-3 (1-1320)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              :::   :::|||::::::::::   :::||||||       |||   |||:::   :::
Db        934 AAAGAACAACAACAGAAGCAGAAACAGGAAGAGGAACCTTTGGAAGTGCAAAGGTACAGA 993

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              |||   ||||||      ||||||||||||   ||||||      |||      |||
Db        994 GCTGAATTGTCTGAAGACGATGTATTTGTAATTCCAGCA------GCTTATCCATTTGTC 1047

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
                 :::             ||||||   :::|||||||||:::||||||:::|||||||||
Db       1048 GTCAACGCTACCTCAAACCTCAATTTCCTTGCTTTTGGTATCAATGCTGAGAACAACCAG 1107

Qy         61 ArgIlePheLeuAlaGlyArgProPheLeuAsnHisLysGlnAsnThrAsnValIle 80
              |||   ||||||||||||                    :::      ||||||:::
Db       1108 AGGAACTTCCTTGCAGGC----------------------GAGAAAGACAATGTGGTA 1143

Qy         81 Lys 81
              :::
Db       1144 AGG 1146

RESULT 14
US-09-108-010B-1
; Sequence 1, Application US/09108010B
; Patent No. 6362399
;   GENERAL INFORMATION:
;         APPLICANT: ANTHONY JOHN KINNEY
;                    GARY MICHAEL FADER
;         TITLE OF INVENTION: SUPPRESSION OF SPECIFIC
;                             CLASSES OF SOYBEAN SEED
;                             PROTEIN GENES
;         NUMBER OF SEQUENCES: 22
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
;              STREET: 1007 MARKET STREET
;              CITY: WILMINGTON
;              STATE: DELAWARE
;              COUNTRY: UNITED STATES OF AMERICA
;              ZIP: 19898
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: DISKETTE, 3.50 INCH
;              COMPUTER: IBM PC COMPATIBLE
;              OPERATING SYSTEM: MICROSOFT WINDOWS 95
;              SOFTWARE: MICROFSOFT WORD 7.0A
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/09/108,010B
;              FILING DATE: 30-Jun-1998
;              CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 60/019,940
;              FILING DATE: JUNE 14,1996
;         ATTORNEY/AGENT INFORMATION:
;              NAME: LYNNE M. CHRISTENBURY
;              REGISTRATION NUMBER: 30,971
;              REFERENCE/DOCKET NUMBER: BB-1071-A
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: 302-992-5481
;              TELEFAX: 302-773-0164
;    INFORMATION FOR SEQ ID NO: 1:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 1818 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: cDNA
;         SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-108-010B-1

Alignment Scores:
Pred. No.:              4.16e-10        Length:         1818
Score:                  137.00          Matches:        30
Percent Similarity:     66.67%          Conservative:   14
Best Local Similarity:  45.45%          Mismatches:     20
Query Match:            29.91%          Indels:         2
DB:                     3               Gaps:           1

US-10-691-590-1 (1-89) x US-09-108-010B-1 (1-1818)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              :::   :::|||:::|||:::   :::||||||       |||   ::::::   :::
Db       1432 AAAGAACAACAACAGGAGCAGCAACAGGAAGAGCAACCTTTGGAAGTGCGGAAATATAGA 1491

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              |||   ||||||      |||:::||||||   |||||||||:::   |||
Db       1492 GCCGAATTGTCTGAACAAGATATATTTGTAATCCCAGCAGGTTATCCAGTTGTGGTCAAC 1551

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              |||:::   :::      ||||||      |||   |||:::||||||:::|||||||||
Db       1552 GCTACCTCAAAT------CTGAATTTCTTTGCTATTGGTATTAATGCCGAGAACAACCAG 1605

Qy         61 ArgIlePheLeuAlaGly 66
              |||   ||||||||||||
Db       1606 AGGAACTTCCTCGCAGGT 1623

RESULT 15
US-09-758-652-1
; Sequence 1, Application US/09758652
; Patent No. 6703544
;   GENERAL INFORMATION:
;         APPLICANT: ANTHONY JOHN KINNEY
;                    GARY MICHAEL FADER
;         TITLE OF INVENTION: SUPPRESSION OF SPECIFIC
;                             CLASSES OF SOYBEAN SEED
;                             PROTEIN GENES
;         NUMBER OF SEQUENCES: 21
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
;              STREET: 1007 MARKET STREET
;              CITY: WILMINGTON
;              STATE: DELAWARE
;              COUNTRY: UNITED STATES OF AMERICA
```

; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROFSOFT WORD 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/758,652
; FILING DATE: 11-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/019,940
; FILING DATE: JUNE 14,1996
; ATTORNEY/AGENT INFORMATION:
; NAME: LYNNE M. CHRISTENBURY
; REGISTRATION NUMBER: 30,971
; REFERENCE/DOCKET NUMBER: BB-1071-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-5481
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1818 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-758-652-1

Alignment Scores:

Pred. No.:	4.16e-10	Length:	1818
Score:	137.00	Matches:	30
Percent Similarity:	66.67%	Conservative:	14
Best Local Similarity:	45.45%	Mismatches:	20
Query Match:	29.91%	Indels:	2
DB:	4	Gaps:	1

US-10-691-590-1 (1-89) x US-09-758-652-1 (1-1818)

```
Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              :::   :::|||:::|||:::   :::||||||      |||   ::::::   :::
Db       1432 AAAGAACAACAACAGGAGCAGCAACAGGAAGAGCAACCTTTGGAAGTGCGGAAATATAGA 1491

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              |||   ||||||      |||:::||||||   ||||||||||:::   |||
Db       1492 GCCGAATTGTCTGAACAAGATATATTTGTAATCCCAGCAGGTTATCCAGTTGTGGTCAAC 1551

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              |||:::   :::      ||||||      |||   |||:::||||||:::|||||||||
Db       1552 GCTACCTCAAAT------CTGAATTTCTTTGCTATTGGTATTAATGCCGAGAACAACCAG 1605

Qy         61 ArgIlePheLeuAlaGly 66
              |||   ||||||||||||
Db       1606 AGGAACTTCCTCGCAGGT 1623
```

Search completed: November 5, 2005, 06:01:14
Job time : 201.704 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 5, 2005, 01:54:58 ; Search time 3520.53 Seconds
(without alignments)
962.275 Million cell updates/sec

Title: US-10-691-590-1
Perfect score: 458
Sequence: 1 RREQEEESEEETFGEFQQVK.....FLNHKQNTNVIKFTVKASAY 89

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10691590/runat_02112005_120022_13171/app_query.fasta_1.462
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10691590_@CGN_1_1_4385_@runat_02112005_120022_13171 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	339	74.0	338	6	CA795230	CA795230 Cac_BL_22
2	339	74.0	359	6	CA796844	CA796844 Cac_BL_38
3	339	74.0	487	6	CA794382	CA794382 Cac_BL_16
4	339	74.0	498	6	CA794271	CA794271 Cac_BL_17
5	339	74.0	498	6	CA794288	CA794288 Cac_BL_17
6	339	74.0	499	6	CA794282	CA794282 Cac_BL_17
7	339	74.0	500	6	CA797354	CA797354 Cac_BL_44
8	339	74.0	505	6	CA795194	CA795194 Cac_BL_22
9	339	74.0	505	6	CA796872	CA796872 Cac_BL_39
10	339	74.0	508	6	CA795155	CA795155 Cac_BL_21
11	339	74.0	510	6	CA795157	CA795157 Cac_BL_21
12	339	74.0	510	6	CA797074	CA797074 Cac_BL_41
13	339	74.0	510	6	CA797130	CA797130 Cac_BL_42
14	339	74.0	511	6	CA795185	CA795185 Cac_BL_22
15	339	74.0	512	6	CA797107	CA797107 Cac_BL_41
16	339	74.0	514	6	CA796326	CA796326 Cac_BL_33
17	339	74.0	517	6	CA795470	CA795470 Cac_BL_25
18	339	74.0	517	6	CA797325	CA797325 Cac_BL_44
19	339	74.0	524	6	CA794505	CA794505 Cac_BL_15
20	339	74.0	524	6	CA796851	CA796851 Cac_BL_38
21	339	74.0	526	6	CA795003	CA795003 Cac_BL_20
22	339	74.0	526	6	CA795423	CA795423 Cac_BL_24
23	339	74.0	528	6	CA794978	CA794978 Cac_BL_20
24	339	74.0	528	6	CA796838	CA796838 Cac_BL_38
25	339	74.0	528	6	CA796893	CA796893 Cac_BL_39
26	339	74.0	529	6	CA794376	CA794376 Cac_BL_16
27	339	74.0	530	6	CA796006	CA796006 Cac_BL_30
28	339	74.0	530	6	CA797179	CA797179 Cac_BL_42
29	339	74.0	533	6	CA794366	CA794366 Cac_BL_16
30	339	74.0	534	6	CA796817	CA796817 Cac_BL_38
31	339	74.0	534	6	CA797168	CA797168 Cac_BL_42
32	339	74.0	536	6	CA794760	CA794760 Cac_BL_18
33	339	74.0	536	6	CA796191	CA796191 Cac_BL_32
34	339	74.0	537	6	CA794789	CA794789 Cac_BL_18
35	339	74.0	537	6	CA797261	CA797261 Cac_BL_43
36	339	74.0	537	6	CA797571	CA797571 Cac_BL_46
37	339	74.0	538	6	CA797162	CA797162 Cac_BL_42
38	339	74.0	539	6	CA794360	CA794360 Cac_BL_16
39	339	74.0	539	6	CA794367	CA794367 Cac_BL_16
40	339	74.0	539	6	CA794934	CA794934 Cac_BL_19
41	339	74.0	540	6	CA797170	CA797170 Cac_BL_42
42	339	74.0	540	6	CA797603	CA797603 Cac_BL_46
43	339	74.0	541	6	CA796465	CA796465 Cac_BL_34
44	339	74.0	541	6	CA797248	CA797248 Cac_BL_43
45	339	74.0	544	6	CA796176	CA796176 Cac_BL_31

ALIGNMENTS

RESULT 1
CA795230
LOCUS CA795230 338 bp mRNA linear EST 05-DEC-2002
DEFINITION Cac_BL_226 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
Theobroma cacao cDNA clone Cac_BL_226 5', mRNA sequence.
ACCESSION CA795230
VERSION CA795230.1 GI:26052306
KEYWORDS EST.
SOURCE Theobroma cacao (cacao)
ORGANISM Theobroma cacao
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae; Theobroma.
REFERENCE 1 (bases 1 to 338)
AUTHORS Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D., Retzel,E.R. and Jones,C.A.
TITLE Gene discovery and microarray analysis of cacao (Theobroma cacao L.) varieties
JOURNAL Planta 216 (2), 255-264 (2002)
MEDLINE 22337596
PUBMED 12447539
COMMENT Contact: Jones, Paul
Masterfoods
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
Tel: +44 1664 416644
Email: Paul.Jones@eu.effem.com
Seq primer: T3.
FEATURES Location/Qualifiers
source 1. .338
/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"

```
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_226"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:              3.8e-36        Length:         338
Score:                  339.00         Matches:        66
Percent Similarity:     100.00%        Conservative:   1
Best Local Similarity:  98.51%         Mismatches:     0
Query Match:            74.02%         Indels:         0
DB:                     6              Gaps:           0

US-10-691-590-1 (1-89) x CA795230 (1-338)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         93 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 152

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        153 GCCCCATTGTCACCTGGTGACGTCTTTGTAGCCCCGGCAGGCCATGCAGTTACATTCTTT 212

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 GCATCCAAAGACCAGCCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAGAACAACCAG 272

Qy         61 ArgIlePheLeuAlaGlyArg 67
              ||||||||||||||||||:::
Db        273 AGAATTTTCCTTGCAGGGAAA 293

RESULT 2
CA796844
LOCUS       CA796844                359 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_3891 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_3891 5', mRNA sequence.
ACCESSION   CA796844
VERSION     CA796844.1  GI:26053930
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 359)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .359
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_3891"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:              4.11e-36       Length:         359
Score:                  339.00         Matches:        66
Percent Similarity:     100.00%        Conservative:   1
Best Local Similarity:  98.51%         Mismatches:     0
Query Match:            74.02%         Indels:         0
DB:                     6              Gaps:           0

US-10-691-590-1 (1-89) x CA796844 (1-359)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         90 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 149

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        150 GCCCCATTGTCACCTGGTGACGTCTTTGTAGCCCCGGCAGGCCATGCAGTTACATTCTTT 209

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        210 GCATCCAAAGACCAGCCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAGAACAACCAG 269

Qy         61 ArgIlePheLeuAlaGlyArg 67
              ||||||||||||||||||:::
Db        270 AGAATTTTCCTTGCAGGGAAA 290

RESULT 3
CA794382
LOCUS       CA794382                487 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_1657 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_1657 5', mRNA sequence.
ACCESSION   CA794382
VERSION     CA794382.1  GI:26051458
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_1657"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
```

```
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:                6.12e-36        Length:         487
Score:                    339.00          Matches:        66
Percent Similarity:       100.00%         Conservative:   1
Best Local Similarity:    98.51%          Mismatches:     0
Query Match:              74.02%          Indels:         0
DB:                       6               Gaps:           0

US-10-691-590-1 (1-89) x CA794382 (1-487)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        170 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 229

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        230 GCCCCATTGTCACCTGGTGACGTCTTTGTAGCCCCGGCAGGCCATGCAGTTACATTCTTT 289

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        290 GCATCCAAAGACCAGCCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAGAACAACCAG 349

Qy         61 ArgIlePheLeuAlaGlyArg 67
              |||||||||||||||||:::
Db        350 AGAATTTTCCTTGCAGGGAAA 370

RESULT 4
CA794271
LOCUS       CA794271                498 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_1751 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_1751 5', mRNA sequence.
ACCESSION   CA794271
VERSION     CA794271.1  GI:26051347
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .498
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_1751"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:                6.3e-36         Length:         498
Score:                    339.00          Matches:        66
Percent Similarity:       100.00%         Conservative:   1
Best Local Similarity:    98.51%          Mismatches:     0
Query Match:              74.02%          Indels:         0
DB:                       6               Gaps:           0

US-10-691-590-1 (1-89) x CA794271 (1-498)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 244

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 GCCCCATTGTCACCTGGTGACGTCTTTGTAGCCCCGGCAGGCCATGCAGTTACATTCTTT 304

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 GCATCCAAAGACCAGCCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAGAACAACCAG 364

Qy         61 ArgIlePheLeuAlaGlyArg 67
              |||||||||||||||||:::
Db        365 AGAATTTTCCTTGCAGGGAAA 385

RESULT 5
CA794288
LOCUS       CA794288                498 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_1724 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_1724 5', mRNA sequence.
ACCESSION   CA794288
VERSION     CA794288.1  GI:26051364
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .498
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_1724"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
```

```
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:                6.3e-36        Length:         498
Score:                    339.00         Matches:        66
Percent Similarity:       100.00%        Conservative:   1
Best Local Similarity:    98.51%         Mismatches:     0
Query Match:              74.02%         Indels:         0
DB:                       6              Gaps:           0

US-10-691-590-1 (1-89) x CA794288 (1-498)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         72 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 131

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 GCCCCATTGTCACCTGGTGACGTCTTTGTAGCCCCGGCAGGCCATGCAGTTACATTCTTT 191

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 GCATCCAAAGACCAGCCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAGAACAACCAG 251

Qy         61 ArgIlePheLeuAlaGlyArg 67
              ||||||||||||||||||:::
Db        252 AGAATTTTCCTTGCAGGGAAA 272

RESULT 6
CA794282
LOCUS       CA794282                 499 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_1763 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_1763 5', mRNA sequence.
ACCESSION   CA794282
VERSION     CA794282.1  GI:26051358
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_1763"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:                6.32e-36       Length:         499
Score:                    339.00         Matches:        66
Percent Similarity:       100.00%        Conservative:   1
Best Local Similarity:    98.51%         Mismatches:     0
Query Match:              74.02%         Indels:         0
DB:                       6              Gaps:           0

US-10-691-590-1 (1-89) x CA794282 (1-499)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 305

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 GCCCCATTGTCACCTGGTGACGTCTTTGTAGCCCCGGCAGGCCATGCAGTTACATTCTTT 365

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 GCATCCAAAGACCAGCCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAGAACAACCAG 425

Qy         61 ArgIlePheLeuAlaGlyArg 67
              ||||||||||||||||||:::
Db        426 AGAATTTTCCTTGCAGGGAAA 446

RESULT 7
CA797354
LOCUS       CA797354                 500 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_4444 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_4444 5', mRNA sequence.
ACCESSION   CA797354
VERSION     CA797354.1  GI:26054440
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_4444"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN
```

```
Alignment Scores:
Pred. No.:                6.33e-36        Length:          500
Score:                    339.00          Matches:         66
Percent Similarity:       100.00%         Conservative:    1
Best Local Similarity:    98.51%          Mismatches:      0
Query Match:              74.02%          Indels:          0
DB:                       6               Gaps:            0

US-10-691-590-1 (1-89) x CA797354 (1-500)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        110 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 169

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        170 GCCCCATTGTCACCTGGTGACGTCTTTGTAGCCCCGGCAGGCCATGCAGTTACATTCTTT 229

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        230 GCATCCAAAGACCAGCCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAGAACAACCAG 289

Qy         61 ArgIlePheLeuAlaGlyArg 67
              ||||||||||||||||:::
Db        290 AGAATTTTCCTTGCAGGGAAA 310

RESULT 8
CA795194
LOCUS       CA795194                505 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_2226 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_2226 5', mRNA sequence.
ACCESSION   CA795194
VERSION     CA795194.1  GI:26052270
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 505)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .505
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_2226"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:                6.41e-36        Length:          505
Score:                    339.00          Matches:         66
Percent Similarity:       100.00%         Conservative:    1
Best Local Similarity:    98.51%          Mismatches:      0
Query Match:              74.02%          Indels:          0
DB:                       6               Gaps:            0

US-10-691-590-1 (1-89) x CA795194 (1-505)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 121

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 GCCCCATTGTCACCTGGTGACGTCTTTGTAGCCCCGGCAGGCCATGCAGTTACATTCTTT 181

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 GCATCCAAAGACCAGCCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAGAACAACCAG 241

Qy         61 ArgIlePheLeuAlaGlyArg 67
              ||||||||||||||||||:::
Db        242 AGAATTTTCCTTGCAGGGAAA 262

RESULT 9
CA796872
LOCUS       CA796872                505 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_3924 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_3924 5', mRNA sequence.
ACCESSION   CA796872
VERSION     CA796872.1  GI:26053958
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 505)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .505
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_3924"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:                6.41e-36        Length:          505
Score:                    339.00          Matches:         66
Percent Similarity:       100.00%         Conservative:    1
```

```
Best Local Similarity:  98.51%          Mismatches:     0
Query Match:            74.02%          Indels:         0
DB:                     6               Gaps:           0

US-10-691-590-1 (1-89) x CA796872 (1-505)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 100

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GCCCCATTGTCACCTGGTGACGTCTTTGTAGCCCCGGCAGGCCATGCAGTTACATTCTTT 160

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GCATCCAAAGACCAGCCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAGAACAACCAG 220

Qy         61 ArgIlePheLeuAlaGlyArg 67
              |||||||||||||||||:::
Db        221 AGAATTTTCCTTGCAGGGAAA 241

RESULT 10
CA795155
LOCUS       CA795155                508 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_2191 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_2191 5', mRNA sequence.
ACCESSION   CA795155
VERSION     CA795155.1  GI:26052231
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 508)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .508
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_2191"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:              6.46e-36        Length:         508
Score:                  339.00          Matches:        66
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  98.51%          Mismatches:     0
Query Match:            74.02%          Indels:         0
DB:                     6               Gaps:           0

US-10-691-590-1 (1-89) x CA795155 (1-508)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 100

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GCCCCATTGTCACCTGGTGACGTCTTTGTAGCCCCGGCAGGCCATGCAGTTACATTCTTT 160

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GCATCCAAAGACCAGCCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAGAACAACCAG 220

Qy         61 ArgIlePheLeuAlaGlyArg 67
              |||||||||||||||||:::
Db        221 AGAATTTTCCTTGCAGGGAAA 241

RESULT 11
CA795157
LOCUS       CA795157                510 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_2193 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_2193 5', mRNA sequence.
ACCESSION   CA795157
VERSION     CA795157.1  GI:26052233
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 510)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .510
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_2193"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:              6.5e-36         Length:         510
Score:                  339.00          Matches:        66
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  98.51%          Mismatches:     0
Query Match:            74.02%          Indels:         0
DB:                     6               Gaps:           0
```

```
US-10-691-590-1 (1-89) x CA795157 (1-510)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         72 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 131

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 GCCCCATTGTCACCTGGTGACGTCTTTGTAGCCCCGGCAGGCCATGCAGTTACATTCTTT 191

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 GCATCCAAAGACCAGCCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAGAACAACCAG 251

Qy         61 ArgIlePheLeuAlaGlyArg 67
              ||||||||||||||||||:::
Db        252 AGAATTTTCCTTGCAGGGAAA 272

RESULT 12
CA797074
LOCUS       CA797074                510 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_4159 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_4159 5', mRNA sequence.
ACCESSION   CA797074
VERSION     CA797074.1  GI:26054160
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 510)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .510
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_4159"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:              6.5e-36         Length:         510
Score:                  339.00          Matches:        66
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  98.51%          Mismatches:     0
Query Match:            74.02%          Indels:         0
DB:                     6               Gaps:           0

US-10-691-590-1 (1-89) x CA797074 (1-510)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 365

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 GCCCCATTGTCACCTGGTGACGTCTTTGTAGCCCCGGCAGGCCATGCAGTTACATTCTTT 425

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 GCATCCAAAGACCAGCCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAGAACAACCAG 485

Qy         61 ArgIlePheLeuAlaGlyArg 67
              ||||||||||||||||||:::
Db        486 AGAATTTTCCTTGCAGGGAAA 506

RESULT 13
CA797130
LOCUS       CA797130                510 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_4215 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_4215 5', mRNA sequence.
ACCESSION   CA797130
VERSION     CA797130.1  GI:26054216
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 510)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .510
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_4215"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:              6.5e-36         Length:         510
Score:                  339.00          Matches:        66
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  98.51%          Mismatches:     0
Query Match:            74.02%          Indels:         0
DB:                     6               Gaps:           0

US-10-691-590-1 (1-89) x CA797130 (1-510)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        282 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 341

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        342 GCCCCATTGTCACCTGGTGACGTCTTTGTAGCCCCGGCAGGCCATGCAGTTACATTCTTT 401

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        402 GCATCCAAAGACCAGCCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAGAACAACCAG 461

Qy         61 ArgIlePheLeuAlaGlyArg 67
              ||||||||||||||||||:::
Db        462 AGAATTTTCCTTGCAGGGAAA 482

RESULT 14
CA795185
LOCUS       CA795185                511 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_2218 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_2218 5', mRNA sequence.
ACCESSION   CA795185
VERSION     CA795185.1  GI:26052261
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_2218"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:              6.51e-36        Length:         511
Score:                  339.00          Matches:        66
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  98.51%          Mismatches:     0
Query Match:            74.02%          Indels:         0
DB:                     6               Gaps:           0

US-10-691-590-1 (1-89) x CA795185 (1-511)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        271 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 330

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        331 GCCCCATTGTCACCTGGTGACGTCTTTGTAGCCCCGGCAGGCCATGCAGTTACATTCTTT 390

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        391 GCATCCAAAGACCAGCCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAGAACAACCAG 450

Qy         61 ArgIlePheLeuAlaGlyArg 67
              ||||||||||||||||||:::
Db        451 AGAATTTTCCTTGCAGGGAAA 471

RESULT 15
CA797107
LOCUS       CA797107                512 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_4193 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_4193 5', mRNA sequence.
ACCESSION   CA797107
VERSION     CA797107.1  GI:26054193
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_4193"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:              6.53e-36        Length:         512
Score:                  339.00          Matches:        66
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  98.51%          Mismatches:     0
Query Match:            74.02%          Indels:         0
DB:                     6               Gaps:           0

US-10-691-590-1 (1-89) x CA797107 (1-512)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 286

Qy         21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        287 GCCCCATTGTCACCTGGTGACGTCTTTGTAGCCCCGGCAGGCCATGCAGTTACATTCTTT 346

Qy         41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        347 GCATCCAAAGACCAGCCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAGAACAACCAG 406

Qy         61 ArgIlePheLeuAlaGlyArg 67
              ||||||||||||||||||:::
Db        407 AGAATTTTCCTTGCAGGGAAA 427
```

Search completed: November 5, 2005, 05:56:53
Job time : 3527.53 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 5, 2005, 01:54:58 ; Search time 1028.47 Seconds
(without alignments)
962.275 Million cell updates/sec

Title: US-10-691-590-3
Perfect score: 122
Sequence: 1 RREQEEESEEETFGEFXQVXAPLXPG 26

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10691590/runat_02112005_120022_13171/app_query.fasta_1.462
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10691590_@CGN_1_1_4385_@runat_02112005_120022_13171 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	117	95.9	256	6	CA794775	CA794775 Cac_BL_18
2	117	95.9	274	6	CA797731	CA797731 Cac_BL_48
3	117	95.9	338	6	CA795230	CA795230 Cac_BL_22
4	117	95.9	351	6	CA795596	CA795596 Cac_BL_26
5	117	95.9	359	6	CA796844	CA796844 Cac_BL_38
6	117	95.9	362	6	CA796065	CA796065 Cac_BL_30
7	117	95.9	364	6	CA797330	CA797330 Cac_BL_44
8	117	95.9	378	6	CA795334	CA795334 Cac_BL_23
9	117	95.9	423	6	CA796843	CA796843 Cac_BL_38
10	117	95.9	430	6	CA794888	CA794888 Cac_BL_19
11	117	95.9	441	6	CA797867	CA797867 Cac_BL_51
12	117	95.9	463	6	CA795681	CA795681 Cac_BL_27
13	117	95.9	465	6	CA794398	CA794398 Cac_BL_16
14	117	95.9	468	6	CA797010	CA797010 Cac_BL_40
15	117	95.9	474	6	CA795491	CA795491 Cac_BL_25
16	117	95.9	475	6	CA796772	CA796772 Cac_BL_38
17	117	95.9	477	6	CA794451	CA794451 Cac_BL_15
18	117	95.9	482	6	CA795961	CA795961 Cac_BL_29
19	117	95.9	487	6	CA794382	CA794382 Cac_BL_16
20	117	95.9	490	6	CA796238	CA796238 Cac_BL_32
21	117	95.9	491	6	CA797417	CA797417 Cac_BL_45
22	117	95.9	495	6	CA794315	CA794315 Cac_BL_17
23	117	95.9	498	6	CA794271	CA794271 Cac_BL_17
24	117	95.9	498	6	CA794288	CA794288 Cac_BL_17
25	117	95.9	498	6	CA794445	CA794445 Cac_BL_15
26	117	95.9	498	6	CA797323	CA797323 Cac_BL_44
27	117	95.9	499	6	CA794282	CA794282 Cac_BL_17
28	117	95.9	499	6	CA794340	CA794340 Cac_BL_16
29	117	95.9	499	6	CA794715	CA794715 Cac_BL_10
30	117	95.9	500	6	CA794343	CA794343 Cac_BL_16
31	117	95.9	500	6	CA797354	CA797354 Cac_BL_44
32	117	95.9	500	6	CA798279	CA798279 Cac_BL_60
33	117	95.9	503	6	CA797220	CA797220 Cac_BL_43
34	117	95.9	505	6	CA795194	CA795194 Cac_BL_22
35	117	95.9	505	6	CA796872	CA796872 Cac_BL_39
36	117	95.9	508	6	CA794856	CA794856 Cac_BL_19
37	117	95.9	508	6	CA795155	CA795155 Cac_BL_21
38	117	95.9	508	6	CA795218	CA795218 Cac_BL_22
39	117	95.9	508	6	CA797423	CA797423 Cac_BL_45
40	117	95.9	509	6	CA797123	CA797123 Cac_BL_42
41	117	95.9	510	6	CA795157	CA795157 Cac_BL_21
42	117	95.9	510	6	CA797074	CA797074 Cac_BL_41
43	117	95.9	510	6	CA797130	CA797130 Cac_BL_42
44	117	95.9	511	6	CA795185	CA795185 Cac_BL_22
45	117	95.9	512	6	CA797107	CA797107 Cac_BL_41

ALIGNMENTS

RESULT 1
CA794775
LOCUS CA794775 256 bp mRNA linear EST 05-DEC-2002
DEFINITION Cac_BL_1827 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
Theobroma cacao cDNA clone Cac_BL_1827 5', mRNA sequence.
ACCESSION CA794775
VERSION CA794775.1 GI:26051851
KEYWORDS EST.
SOURCE Theobroma cacao (cacao)
ORGANISM Theobroma cacao
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae; Theobroma.
REFERENCE 1 (bases 1 to 256)
AUTHORS Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D., Retzel,E.R. and Jones,C.A.
TITLE Gene discovery and microarray analysis of cacao (Theobroma cacao L.) varieties
JOURNAL Planta 216 (2), 255-264 (2002)
MEDLINE 22337596
PUBMED 12447539
COMMENT Contact: Jones, Paul
Masterfoods
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
Tel: +44 1664 416644
Email: Paul.Jones@eu.effem.com
Seq primer: T3.
FEATURES Location/Qualifiers
source 1. .256
/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"

/db_xref="taxon:3641"
/clone="Cac_BL_1827"
/tissue_type="Mature leaf and mature bean"
/cell_type="Whole organ"
/dev_stage="maturity"
/lab_host="XL-1 Blue MRF'"
/clone_lib="Cac_BL (Bean and Leaf from Amelonardo type Cacao)"
/note="Vector: pBK-CMV; Bean and leaf tissue from an Amelonado type Cacao tree."

ORIGIN

Alignment Scores:

Pred. No.:	3.13e-09	Length:	256
Score:	117.00	Matches:	23
Percent Similarity:	88.46%	Conservative:	0
Best Local Similarity:	88.46%	Mismatches:	3
Query Match:	95.90%	Indels:	0
DB:	6	Gaps:	0

US-10-691-590-3 (1-26) x CA794775 (1-256)

```
Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        105 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 164

Qy         21 AlaProLeu***ProGly 26
              |||||||||   ||||||
Db        165 GCCCCATTGTCACCTGGT 182
```

RESULT 2
CA797731
LOCUS CA797731 274 bp mRNA linear EST 05-DEC-2002
DEFINITION Cac_BL_483 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
Theobroma cacao cDNA clone Cac_BL_483 5', mRNA sequence.
ACCESSION CA797731
VERSION CA797731.1 GI:26054817
KEYWORDS EST.
SOURCE Theobroma cacao (cacao)
ORGANISM Theobroma cacao
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae; Theobroma.
REFERENCE 1 (bases 1 to 274)
AUTHORS Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D., Retzel,E.R. and Jones,C.A.
TITLE Gene discovery and microarray analysis of cacao (Theobroma cacao L.) varieties
JOURNAL Planta 216 (2), 255-264 (2002)
MEDLINE 22337596
PUBMED 12447539
COMMENT Contact: Jones, Paul
Masterfoods
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
Tel: +44 1664 416644
Email: Paul.Jones@eu.effem.com
Seq primer: T3.
FEATURES Location/Qualifiers
source 1. .274
/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"
/db_xref="taxon:3641"
/clone="Cac_BL_483"
/tissue_type="Mature leaf and mature bean"
/cell_type="Whole organ"
/dev_stage="maturity"
/lab_host="XL-1 Blue MRF'"
/clone_lib="Cac_BL (Bean and Leaf from Amelonardo type Cacao)"
/note="Vector: pBK-CMV; Bean and leaf tissue from an Amelonado type Cacao tree."

ORIGIN

Alignment Scores:

Pred. No.:	3.38e-09	Length:	274
Score:	117.00	Matches:	23
Percent Similarity:	88.46%	Conservative:	0
Best Local Similarity:	88.46%	Mismatches:	3
Query Match:	95.90%	Indels:	0
DB:	6	Gaps:	0

US-10-691-590-3 (1-26) x CA797731 (1-274)

```
Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db         63 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 122

Qy         21 AlaProLeu***ProGly 26
              |||||||||   ||||||
Db        123 GCCCCATTGTCACCTGGT 140
```

RESULT 3
CA795230
LOCUS CA795230 338 bp mRNA linear EST 05-DEC-2002
DEFINITION Cac_BL_226 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
Theobroma cacao cDNA clone Cac_BL_226 5', mRNA sequence.
ACCESSION CA795230
VERSION CA795230.1 GI:26052306
KEYWORDS EST.
SOURCE Theobroma cacao (cacao)
ORGANISM Theobroma cacao
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae; Theobroma.
REFERENCE 1 (bases 1 to 338)
AUTHORS Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D., Retzel,E.R. and Jones,C.A.
TITLE Gene discovery and microarray analysis of cacao (Theobroma cacao L.) varieties
JOURNAL Planta 216 (2), 255-264 (2002)
MEDLINE 22337596
PUBMED 12447539
COMMENT Contact: Jones, Paul
Masterfoods
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
Tel: +44 1664 416644
Email: Paul.Jones@eu.effem.com
Seq primer: T3.
FEATURES Location/Qualifiers
source 1. .338
/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"
/db_xref="taxon:3641"
/clone="Cac_BL_226"
/tissue_type="Mature leaf and mature bean"
/cell_type="Whole organ"
/dev_stage="maturity"
/lab_host="XL-1 Blue MRF'"
/clone_lib="Cac_BL (Bean and Leaf from Amelonardo type Cacao)"
/note="Vector: pBK-CMV; Bean and leaf tissue from an Amelonado type Cacao tree."

ORIGIN

Alignment Scores:

Pred. No.:	4.26e-09	Length:	338
Score:	117.00	Matches:	23
Percent Similarity:	88.46%	Conservative:	0
Best Local Similarity:	88.46%	Mismatches:	3
Query Match:	95.90%	Indels:	0
DB:	6	Gaps:	0

```
US-10-691-590-3 (1-26) x CA795230 (1-338)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db         93 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 152

Qy         21 AlaProLeu***ProGly 26
              |||||||||   ||||||
Db        153 GCCCCATTGTCACCTGGT 170

RESULT 4
CA795596
LOCUS       CA795596                351 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_2623 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_2623 5', mRNA sequence.
ACCESSION   CA795596
VERSION     CA795596.1  GI:26052672
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 351)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .351
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_2623"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:              4.44e-09        Length:         351
Score:                  117.00          Matches:        23
Percent Similarity:     88.46%          Conservative:   0
Best Local Similarity:  88.46%          Mismatches:     3
Query Match:            95.90%          Indels:         0
DB:                     6               Gaps:           0

US-10-691-590-3 (1-26) x CA795596 (1-351)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        239 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 298

Qy         21 AlaProLeu***ProGly 26
              |||||||||   ||||||
Db        299 GCCCCATTGTCACCTGGG 316

RESULT 5
CA796844
LOCUS       CA796844                359 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_3891 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_3891 5', mRNA sequence.
ACCESSION   CA796844
VERSION     CA796844.1  GI:26053930
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 359)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .359
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_3891"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:              4.56e-09        Length:         359
Score:                  117.00          Matches:        23
Percent Similarity:     88.46%          Conservative:   0
Best Local Similarity:  88.46%          Mismatches:     3
Query Match:            95.90%          Indels:         0
DB:                     6               Gaps:           0

US-10-691-590-3 (1-26) x CA796844 (1-359)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db         90 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 149

Qy         21 AlaProLeu***ProGly 26
              |||||||||   ||||||
Db        150 GCCCCATTGTCACCTGGT 167

RESULT 6
CA796065
LOCUS       CA796065                362 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_3076 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_3076 5', mRNA sequence.
ACCESSION   CA796065
VERSION     CA796065.1  GI:26053141
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
```

```
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 362)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .362
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_3076"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:                4.6e-09        Length:          362
Score:                    117.00         Matches:         23
Percent Similarity:       88.46%         Conservative:    0
Best Local Similarity:    88.46%         Mismatches:      3
Query Match:              95.90%         Indels:          0
DB:                       6              Gaps:            0

US-10-691-590-3 (1-26) x CA796065 (1-362)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        171 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 230

Qy         21 AlaProLeu***ProGly 26
              |||||||||   ||||||
Db        231 GCCCCATTGTCACCTGGT 248

RESULT 7
CA797330
LOCUS       CA797330                364 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_442 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_442 5', mRNA sequence.
ACCESSION   CA797330
VERSION     CA797330.1  GI:26054416
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 364)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .364
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_442"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:                4.63e-09       Length:          364
Score:                    117.00         Matches:         23
Percent Similarity:       88.46%         Conservative:    0
Best Local Similarity:    88.46%         Mismatches:      3
Query Match:              95.90%         Indels:          0
DB:                       6              Gaps:            0

US-10-691-590-3 (1-26) x CA797330 (1-364)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        186 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 245

Qy         21 AlaProLeu***ProGly 26
              |||||||||   ||||||
Db        246 GCCCCATTGTCACCTGGT 263

RESULT 8
CA795334
LOCUS       CA795334                378 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_2374 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_2374 5', mRNA sequence.
ACCESSION   CA795334
VERSION     CA795334.1  GI:26052410
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 378)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .378
```

/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"
/db_xref="taxon:3641"
/clone="Cac_BL_2374"
/tissue_type="Mature leaf and mature bean"
/cell_type="Whole organ"
/dev_stage="maturity"
/lab_host="XL-1 Blue MRF'"
/clone_lib="Cac_BL (Bean and Leaf from Amelonardo type Cacao)"
/note="Vector: pBK-CMV; Bean and leaf tissue from an Amelonado type Cacao tree."

ORIGIN

Alignment Scores:

Pred. No.:	4.83e-09	Length:	378
Score:	117.00	Matches:	23
Percent Similarity:	88.46%	Conservative:	0
Best Local Similarity:	88.46%	Mismatches:	3
Query Match:	95.90%	Indels:	0
DB:	6	Gaps:	0

US-10-691-590-3 (1-26) x CA795334 (1-378)

```
Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db        243 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 302

Qy         21 AlaProLeu***ProGly 26
              |||||||||   ||||||
Db        303 GCCCCATTGTCACCTGGT 320
```

RESULT 9
CA796843
LOCUS CA796843 423 bp mRNA linear EST 05-DEC-2002
DEFINITION Cac_BL_3890 Cac_BL (Bean and Leaf from Amelonardo type Cacao) Theobroma cacao cDNA clone Cac_BL_3890 5', mRNA sequence.
ACCESSION CA796843
VERSION CA796843.1 GI:26053929
KEYWORDS EST.
SOURCE Theobroma cacao (cacao)
ORGANISM Theobroma cacao
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae; Theobroma.
REFERENCE 1 (bases 1 to 423)
AUTHORS Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D., Retzel,E.R. and Jones,C.A.
TITLE Gene discovery and microarray analysis of cacao (Theobroma cacao L.) varieties
JOURNAL Planta 216 (2), 255-264 (2002)
MEDLINE 22337596
PUBMED 12447539
COMMENT Contact: Jones, Paul
Masterfoods
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
Tel: +44 1664 416644
Email: Paul.Jones@eu.effem.com
Seq primer: T3.

FEATURES Location/Qualifiers
source 1. .423
/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"
/db_xref="taxon:3641"
/clone="Cac_BL_3890"
/tissue_type="Mature leaf and mature bean"
/cell_type="Whole organ"
/dev_stage="maturity"
/lab_host="XL-1 Blue MRF'"
/clone_lib="Cac_BL (Bean and Leaf from Amelonardo type Cacao)"
/note="Vector: pBK-CMV; Bean and leaf tissue from an Amelonado type Cacao tree."

ORIGIN

Alignment Scores:

Pred. No.:	5.46e-09	Length:	423
Score:	117.00	Matches:	23
Percent Similarity:	88.46%	Conservative:	0
Best Local Similarity:	88.46%	Mismatches:	3
Query Match:	95.90%	Indels:	0
DB:	6	Gaps:	0

US-10-691-590-3 (1-26) x CA796843 (1-423)

```
Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db        269 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 328

Qy         21 AlaProLeu***ProGly 26
              |||||||||   ||||||
Db        329 GCCCCATTGTCACCTGGT 346
```

RESULT 10
CA794888
LOCUS CA794888 430 bp mRNA linear EST 05-DEC-2002
DEFINITION Cac_BL_1941 Cac_BL (Bean and Leaf from Amelonardo type Cacao) Theobroma cacao cDNA clone Cac_BL_1941 5', mRNA sequence.
ACCESSION CA794888
VERSION CA794888.1 GI:26051964
KEYWORDS EST.
SOURCE Theobroma cacao (cacao)
ORGANISM Theobroma cacao
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae; Theobroma.
REFERENCE 1 (bases 1 to 430)
AUTHORS Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D., Retzel,E.R. and Jones,C.A.
TITLE Gene discovery and microarray analysis of cacao (Theobroma cacao L.) varieties
JOURNAL Planta 216 (2), 255-264 (2002)
MEDLINE 22337596
PUBMED 12447539
COMMENT Contact: Jones, Paul
Masterfoods
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
Tel: +44 1664 416644
Email: Paul.Jones@eu.effem.com
Seq primer: T3.

FEATURES Location/Qualifiers
source 1. .430
/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"
/db_xref="taxon:3641"
/clone="Cac_BL_1941"
/tissue_type="Mature leaf and mature bean"
/cell_type="Whole organ"
/dev_stage="maturity"
/lab_host="XL-1 Blue MRF'"
/clone_lib="Cac_BL (Bean and Leaf from Amelonardo type Cacao)"
/note="Vector: pBK-CMV; Bean and leaf tissue from an Amelonado type Cacao tree."

ORIGIN

Alignment Scores:

Pred. No.:	5.57e-09	Length:	430
Score:	117.00	Matches:	23
Percent Similarity:	88.46%	Conservative:	0
Best Local Similarity:	88.46%	Mismatches:	3

```
Query Match:           95.90%        Indels:        0
DB:                    6             Gaps:          0

US-10-691-590-3 (1-26) x CA794888 (1-430)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        271 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 330

Qy         21 AlaProLeu***ProGly 26
              |||||||||   ||||||
Db        331 GCCCCATTGTCACCTGGT 348

RESULT 11
CA797867
LOCUS       CA797867                441 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_51 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_51 5', mRNA sequence.
ACCESSION   CA797867
VERSION     CA797867.1  GI:26054953
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 441)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .441
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_51"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:              5.72e-09      Length:        441
Score:                  117.00        Matches:       23
Percent Similarity:     88.46%        Conservative:  0
Best Local Similarity:  88.46%        Mismatches:    3
Query Match:            95.90%        Indels:        0
DB:                     6             Gaps:          0

US-10-691-590-3 (1-26) x CA797867 (1-441)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        318 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 377

Qy         21 AlaProLeu***ProGly 26
              |||||||||   ||||||
Db        378 GCCCCATTGTCACCTGGT 395

RESULT 12
CA795681
LOCUS       CA795681                463 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_2707 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_2707 5', mRNA sequence.
ACCESSION   CA795681
VERSION     CA795681.1  GI:26052757
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 463)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .463
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_2707"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:              6.04e-09      Length:        463
Score:                  117.00        Matches:       23
Percent Similarity:     88.46%        Conservative:  0
Best Local Similarity:  88.46%        Mismatches:    3
Query Match:            95.90%        Indels:        0
DB:                     6             Gaps:          0

US-10-691-590-3 (1-26) x CA795681 (1-463)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        282 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 341

Qy         21 AlaProLeu***ProGly 26
              |||||||||   ||||||
Db        342 GCCCCATTGTCACCTGGT 359

RESULT 13
CA794398
LOCUS       CA794398                465 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_1631 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_1631 5', mRNA sequence.
ACCESSION   CA794398
VERSION     CA794398.1  GI:26051474
```

```
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 465)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .465
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_1631"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:                6.07e-09        Length:          465
Score:                    117.00          Matches:         23
Percent Similarity:       88.46%          Conservative:    0
Best Local Similarity:    88.46%          Mismatches:      3
Query Match:              95.90%          Indels:          0
DB:                       6               Gaps:            0

US-10-691-590-3 (1-26) x CA794398 (1-465)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        358 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 417

Qy         21 AlaProLeu***ProGly 26
              |||||||||   ||||||
Db        418 GCCCCATTGTCACCTGGT 435

RESULT 14
CA797010
LOCUS       CA797010                468 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_4090 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_4090 5', mRNA sequence.
ACCESSION   CA797010
VERSION     CA797010.1  GI:26054096
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 468)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .468
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_4090"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:                6.11e-09        Length:          468
Score:                    117.00          Matches:         23
Percent Similarity:       88.46%          Conservative:    0
Best Local Similarity:    88.46%          Mismatches:      3
Query Match:              95.90%          Indels:          0
DB:                       6               Gaps:            0

US-10-691-590-3 (1-26) x CA797010 (1-468)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        352 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 411

Qy         21 AlaProLeu***ProGly 26
              |||||||||   ||||||
Db        412 GCCCCATTGTCACCTGGT 429

RESULT 15
CA795491
LOCUS       CA795491                474 bp    mRNA    linear   EST 05-DEC-2002
DEFINITION  Cac_BL_2524 Cac_BL (Bean and Leaf from Amelonardo type Cacao)
            Theobroma cacao cDNA clone Cac_BL_2524 5', mRNA sequence.
ACCESSION   CA795491
VERSION     CA795491.1  GI:26052567
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 474)
  AUTHORS   Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
            Retzel,E.R. and Jones,C.A.
  TITLE     Gene discovery and microarray analysis of cacao (Theobroma cacao
            L.) varieties
  JOURNAL   Planta 216 (2), 255-264 (2002)
  MEDLINE   22337596
   PUBMED   12447539
COMMENT     Contact: Jones, Paul
            Masterfoods
            3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
            Tel: +44 1664 416644
            Email: Paul.Jones@eu.effem.com
```

```
           Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .474
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /strain="Amelonado type"
                     /db_xref="taxon:3641"
                     /clone="Cac_BL_2524"
                     /tissue_type="Mature leaf and mature bean"
                     /cell_type="Whole organ"
                     /dev_stage="maturity"
                     /lab_host="XL-1 Blue MRF'"
                     /clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
                     Cacao)"
                     /note="Vector: pBK-CMV; Bean and leaf tissue from an
                     Amelonado type Cacao tree."
ORIGIN

Alignment Scores:
Pred. No.:              6.2e-09        Length:         474
Score:                  117.00         Matches:        23
Percent Similarity:     88.46%         Conservative:   0
Best Local Similarity:  88.46%         Mismatches:     3
Query Match:            95.90%         Indels:         0
DB:                     6              Gaps:           0

US-10-691-590-3 (1-26) x CA795491 (1-474)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        247 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 306

Qy         21 AlaProLeu***ProGly 26
              |||||||||   ||||||
Db        307 GCCCCATTGTCACCTGGT 324


Search completed: November  5, 2005, 05:56:57
Job time : 1032.47 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 5, 2005, 02:54:27 ; Search time 263.165 Seconds
(without alignments)
817.035 Million cell updates/sec

Title: US-10-691-590-3
Perfect score: 122
Sequence: 1 RREQEEESEEETFGEFXQVXAPLXPG 26

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10691590/runat_02112005_120023_13233/app_query.fasta_1.462
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10691590_@CGN_1_1_582_@runat_02112005_120023_13233
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
23: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
24: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq2:*
25: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
26: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
27: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
28: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	61	50.0	2362	20	US-10-437-963-58448	Sequence 58448, A
c	2	59	48.4	1845	17	US-10-029-386-24791	Sequence 24791, A
	3	59	48.4	2365	21	US-10-723-860-1651	Sequence 1651, Ap
	4	59	48.4	2365	22	US-10-741-600-24	Sequence 24, Appl
	5	59	48.4	2375	22	US-10-741-600-23	Sequence 23, Appl
	6	59	48.4	2382	21	US-10-723-860-6048	Sequence 6048, Ap
c	7	59	48.4	14123	22	US-10-741-600-17921	Sequence 17921, A
	8	59	48.4	16220	22	US-10-741-600-17562	Sequence 17562, A
c	9	59	48.4	43623	22	US-10-741-600-17818	Sequence 17818, A
	10	57	46.7	584	21	US-10-425-115-160535	Sequence 160535,
	11	57	46.7	1047	20	US-10-437-963-60949	Sequence 60949, A
c	12	55	45.1	2589	26	US-11-097-143-11710	Sequence 11710, A
	13	54.5	44.7	4754	9	US-09-982-091A-1	Sequence 1, Appli
	14	54	44.3	575	13	US-09-925-065A-133584	Sequence 133584,
	15	54	44.3	625	13	US-09-925-065A-854915	Sequence 854915,
	16	54	44.3	658	13	US-09-925-065A-93913	Sequence 93913, A
	17	54	44.3	684	20	US-10-437-963-2415	Sequence 2415, Ap
	18	54	44.3	752	19	US-10-424-599-31267	Sequence 31267, A
c	19	54	44.3	11817	16	US-10-156-761-2884	Sequence 2884, Ap
c	20	54	44.3	65253	18	US-10-085-117-331	Sequence 331, App
c	21	54	44.3	125746	16	US-10-156-761-15102	Sequence 15102, A
	22	54	44.3	9025608	16	US-10-156-761-1	Sequence 1, Appli
c	23	53	43.4	316	20	US-10-674-124A-12347	Sequence 12347, A
c	24	53	43.4	727	21	US-10-425-115-90035	Sequence 90035, A
c	25	53	43.4	1017	20	US-10-437-963-45963	Sequence 45963, A
	26	53	43.4	1301	21	US-10-739-930-1674	Sequence 1674, Ap
	27	53	43.4	4756	9	US-09-982-091A-3	Sequence 3, Appli
	28	53	43.4	4804	21	US-10-357-930-30217	Sequence 30217, A
	29	53	43.4	111206	14	US-10-087-192-1957	Sequence 1957, Ap
c	30	53	43.4	134841	14	US-10-087-192-1987	Sequence 1987, Ap
c	31	52.5	43.0	1281	20	US-10-437-963-15584	Sequence 15584, A
	32	52	42.6	403	19	US-10-276-774-259	Sequence 259, App
	33	52	42.6	405	9	US-09-770-423-713	Sequence 713, App
c	34	52	42.6	698	14	US-10-027-632-148019	Sequence 148019,
c	35	52	42.6	698	18	US-10-027-632-148019	Sequence 148019,
	36	52	42.6	1297	19	US-10-425-114-29320	Sequence 29320, A
	37	52	42.6	1540	20	US-10-437-963-5733	Sequence 5733, Ap
	38	52	42.6	1580	19	US-10-425-114-20659	Sequence 20659, A
c	39	52	42.6	1619	19	US-10-424-599-75133	Sequence 75133, A
	40	52	42.6	1874	19	US-10-424-599-75130	Sequence 75130, A
c	41	52	42.6	2667	19	US-10-424-599-91404	Sequence 91404, A
	42	52	42.6	147309	9	US-09-742-312-3	Sequence 3, Appli
	43	52	42.6	147309	17	US-10-436-185-3	Sequence 3, Appli
	44	52	42.6	312477	20	US-10-317-883A-12	Sequence 12, Appl
	45	51.5	42.2	6953	19	US-10-600-230-66	Sequence 66, Appl

ALIGNMENTS

RESULT 1
US-10-437-963-58448/c
; Sequence 58448, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

```
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 58448
;   LENGTH: 2362
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(2362)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_60163C.1
US-10-437-963-58448

Alignment Scores:
Pred. No.:              1.22          Length:        2362
Score:                  61.00         Matches:       12
Percent Similarity:     72.73%        Conservative:  4
Best Local Similarity:  54.55%        Mismatches:    6
Query Match:            50.00%        Indels:        0
DB:                     20            Gaps:          0

US-10-691-590-3 (1-26) x US-10-437-963-58448 (1-2362)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              |||||||||:::||||||:::::::||||||:::|||   |||   |||   |||
Db         66 AGGCGAGAGAGAGAGGAAAAAGCGGAAGAGAAAACAGCGGGGCGATTCGCCCAGCCCGGG 7

Qy         21 AlaPro 22
              |||
Db          6 GATCCA 1

RESULT 2
US-10-029-386-24791/c
; Sequence 24791, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24791
;   LENGTH: 1845
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC008891.4
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =  19
;   OTHER INFORMATION: SWISSPROT HIT: P23327, EVALUE 0.00e+00
;   OTHER INFORMATION: EST_HUMAN HIT: AA778986.1, EVALUE 0.00e+00
;   OTHER INFORMATION: NT HIT: gi11424080, EVALUE 0.00e+00
US-10-029-386-24791

Alignment Scores:
Pred. No.:              2.13          Length:        1845
Score:                  59.00         Matches:       12
Percent Similarity:     64.00%        Conservative:  4
Best Local Similarity:  48.00%        Mismatches:    9
Query Match:            48.36%        Indels:        0
DB:                     17            Gaps:          0

US-10-691-590-3 (1-26) x US-10-029-386-24791 (1-1845)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              :::   |||:::|||||||||   :::||||||         |||      :::|||
Db        442 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 383

Qy         21 AlaProLeu***Pro 25
              |||||||||   |||
Db        382 GCCCCACTGAGCCCA 368

RESULT 3
US-10-723-860-1651
; Sequence 1651, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
;  APPLICANT: Aziz, Natasha
;  APPLICANT:  Ginsburg, Wendy M.
;  APPLICANT:  Zlotnik, Albert
;  TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
;  TITLE OF INVENTION:  Methods for Screening for Soft Tissue Sarcoma Modulators
;  FILE REFERENCE: 05882.0193.NPUS01
;  CURRENT APPLICATION NUMBER: US/10/723,860
;  CURRENT FILING DATE:  2003-11-26
;  PRIOR APPLICATION NUMBER: 60/429,739
;  PRIOR FILING DATE: 2002-11-26
;  NUMBER OF SEQ ID NOS: 8393
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1651
;   LENGTH: 2365
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-723-860-1651

Alignment Scores:
Pred. No.:              2.79          Length:        2365
Score:                  59.00         Matches:       12
Percent Similarity:     64.00%        Conservative:  4
Best Local Similarity:  48.00%        Mismatches:    9
Query Match:            48.36%        Indels:        0
DB:                     21            Gaps:          0

US-10-691-590-3 (1-26) x US-10-723-860-1651 (1-2365)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              :::   |||:::|||||||||   :::||||||         |||      :::|||
Db       1788 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 1847

Qy         21 AlaProLeu***Pro 25
              |||||||||   |||
Db       1848 GCCCCACTGAGCCCA 1862

RESULT 4
US-10-741-600-24
; Sequence 24, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24
;   LENGTH: 2365
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-24

Alignment Scores:
Pred. No.:              2.79          Length:        2365
Score:                  59.00         Matches:       12
Percent Similarity:     64.00%        Conservative:  4
Best Local Similarity:  48.00%        Mismatches:    9
Query Match:            48.36%        Indels:        0
```

```
DB:                    22            Gaps:          0

US-10-691-590-3 (1-26) x US-10-741-600-24 (1-2365)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              :::   |||:::|||||||||   :::||||||         |||      :::|||
Db       1788 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 1847

Qy         21 AlaProLeu***Pro 25
              |||||||||   |||
Db       1848 GCCCCACTGAGCCCA 1862

RESULT 5
US-10-741-600-23
; Sequence 23, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 2375
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-23

Alignment Scores:
Pred. No.:              2.81          Length:        2375
Score:                  59.00         Matches:       12
Percent Similarity:     64.00%        Conservative:  4
Best Local Similarity:  48.00%        Mismatches:    9
Query Match:            48.36%        Indels:        0
DB:                     22            Gaps:          0

US-10-691-590-3 (1-26) x US-10-741-600-23 (1-2375)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              :::   |||:::|||||||||   :::||||||         |||      :::|||
Db       1788 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 1847

Qy         21 AlaProLeu***Pro 25
              |||||||||   |||
Db       1848 GCCCCACTGAGCCCA 1862

RESULT 6
US-10-723-860-6048
; Sequence 6048, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
;  APPLICANT: Aziz, Natasha
;  APPLICANT:  Ginsburg, Wendy M.
;  APPLICANT:  Zlotnik, Albert
;  TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
;  TITLE OF INVENTION:  Methods for Screening for Soft Tissue Sarcoma Modulators
;  FILE REFERENCE: 05882.0193.NPUS01
;  CURRENT APPLICATION NUMBER: US/10/723,860
;  CURRENT FILING DATE:  2003-11-26
;  PRIOR APPLICATION NUMBER: 60/429,739
;  PRIOR FILING DATE: 2002-11-26
;  NUMBER OF SEQ ID NOS: 8393
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6048
;   LENGTH: 2382
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (896)..(984)
;   OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-6048

Alignment Scores:
Pred. No.:              2.82          Length:        2382
Score:                  59.00         Matches:       12
Percent Similarity:     64.00%        Conservative:  4
Best Local Similarity:  48.00%        Mismatches:    9
Query Match:            48.36%        Indels:        0
DB:                     21            Gaps:          0

US-10-691-590-3 (1-26) x US-10-723-860-6048 (1-2382)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              :::   |||:::|||||||||   :::||||||         |||      :::|||
Db       1788 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 1847

Qy         21 AlaProLeu***Pro 25
              |||||||||   |||
Db       1848 GCCCCACTGAGCCCA 1862

RESULT 7
US-10-741-600-17921/c
; Sequence 17921, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17921
;   LENGTH: 14123
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-17921

Alignment Scores:
Pred. No.:              19.4          Length:        14123
Score:                  59.00         Matches:       12
Percent Similarity:     64.00%        Conservative:  4
Best Local Similarity:  48.00%        Mismatches:    9
Query Match:            48.36%        Indels:        0
DB:                     22            Gaps:          0

US-10-691-590-3 (1-26) x US-10-741-600-17921 (1-14123)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              :::   |||:::|||||||||   :::||||||         |||      :::|||
Db       2144 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 2085

Qy         21 AlaProLeu***Pro 25
              |||||||||   |||
Db       2084 GCCCCACTGAGCCCA 2070

RESULT 8
US-10-741-600-17562
; Sequence 17562, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17562
;   LENGTH: 16220
```

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-17562

Alignment Scores:
Pred. No.:              22.6          Length:        16220
Score:                  59.00         Matches:       12
Percent Similarity:     64.00%        Conservative:  4
Best Local Similarity:  48.00%        Mismatches:    9
Query Match:            48.36%        Indels:        0
DB:                     22            Gaps:          0

US-10-691-590-3 (1-26) x US-10-741-600-17562 (1-16220)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              :::   |||:::|||||||||   :::||||||         |||      :::|||
Db       7789 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 7848

Qy         21 AlaProLeu***Pro 25
              |||||||||   |||
Db       7849 GCCCCACTGAGCCCA 7863

RESULT 9
US-10-741-600-17818/c
; Sequence 17818, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17818
;   LENGTH: 43623
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-17818

Alignment Scores:
Pred. No.:              66.1          Length:        43623
Score:                  59.00         Matches:       12
Percent Similarity:     64.00%        Conservative:  4
Best Local Similarity:  48.00%        Mismatches:    9
Query Match:            48.36%        Indels:        0
DB:                     22            Gaps:          0

US-10-691-590-3 (1-26) x US-10-741-600-17818 (1-43623)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              :::   |||:::|||||||||   :::||||||         |||      :::|||
Db      40219 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 40160

Qy         21 AlaProLeu***Pro 25
              |||||||||   |||
Db      40159 GCCCCACTGAGCCCA 40145

RESULT 10
US-10-425-115-160535
; Sequence 160535, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 160535
;   LENGTH: 584
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_77988C.1
US-10-425-115-160535

Alignment Scores:
Pred. No.:              1.4           Length:        584
Score:                  57.00         Matches:       12
Percent Similarity:     59.09%        Conservative:  1
Best Local Similarity:  54.55%        Mismatches:    9
Query Match:            46.72%        Indels:        0
DB:                     21            Gaps:          0

US-10-691-590-3 (1-26) x US-10-425-115-160535 (1-584)

Qy          4 GlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal***AlaProLeu 23
              :::|||||||||   |||||||||      |||||||||            |||   |||
Db        172 GAAGAAGAAGAAGAAGAAGAAGAAAGAGAAGGCGAGTTCGCGGCGGCGATGGCGCAGCTG 231

Qy         24 ***Pro 25
                 |||
Db        232 GCGCCG 237

RESULT 11
US-10-437-963-60949
; Sequence 60949, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 60949
;   LENGTH: 1047
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_6242C.1
US-10-437-963-60949

Alignment Scores:
Pred. No.:              2.64          Length:        1047
Score:                  57.00         Matches:       12
Percent Similarity:     53.85%        Conservative:  2
Best Local Similarity:  46.15%        Mismatches:    12
Query Match:            46.72%        Indels:        0
DB:                     20            Gaps:          0

US-10-691-590-3 (1-26) x US-10-437-963-60949 (1-1047)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||   ||||||:::|||   ||||||                  ||||||
Db        133 CGTCGTCGTCAGGAGGATGAGGAGGAGGAGGTTGTGGCGGCGGCGGCGGCGCAGGTGGCC 192

Qy         21 AlaProLeu***ProGly 26
              :::|||      ||||||
Db        193 TCGCCTCCTTCGCCGGGG 210

RESULT 12
```

```
US-11-097-143-11710/c
; Sequence 11710, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11710
;   LENGTH: 2589
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-11710

Alignment Scores:
Pred. No.:              16.2          Length:          2589
Score:                  55.00         Matches:         12
Percent Similarity:     57.69%        Conservative:    3
Best Local Similarity:  46.15%        Mismatches:      7
Query Match:            45.08%        Indels:          4
DB:                     26            Gaps:            1

US-10-691-590-3 (1-26) x US-11-097-143-11710 (1-2589)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              |||:::|||:::|||   |||   |||   |||                |||:::
Db        578 CGCAAAGAGAGAGAGAGAGAGAGAAAGCGAACGGATG-----------CAAATTGCT 531

Qy         21 AlaProLeu***ProGly 26
              |||||||||   ||||||
Db        530 GCTCCTTTGGCTCCGGGT 513

RESULT 13
US-09-982-091A-1
; Sequence 1, Application US/09982091A
; Patent No. US20020151030A1
; GENERAL INFORMATION:
;  APPLICANT: CALIFORNIA INSTITUTE OF TECHNOLOGY
;  APPLICANT:  KUMAGAI, Akiko
;  APPLICANT:  DUNPHY, William
;  TITLE OF INVENTION: CLASPIN, PROTEINS AND METHODS OF USE THEREOF
;  FILE REFERENCE: CIT1320-1
;  CURRENT APPLICATION NUMBER: US/09/982,091A
;  CURRENT FILING DATE:  2002-10-17
;  PRIOR APPLICATION NUMBER: US 60/241,246
;  PRIOR FILING DATE: 2000-10-17
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 4754
;   TYPE: DNA
;   ORGANISM: Xenopus laevis
US-09-982-091A-1

Alignment Scores:
Pred. No.:              38.5          Length:          4754
Score:                  54.50         Matches:         12
Percent Similarity:     86.67%        Conservative:    1
Best Local Similarity:  80.00%        Mismatches:      1
Query Match:            44.67%        Indels:          1
DB:                     9             Gaps:            1

US-10-691-590-3 (1-26) x US-09-982-091A-1 (1-4754)

Qy          2 ArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe 16
              ||||||||||||||:::   |||||||||   ||||||||||||
Db       2885 AGAGAACAGGAAGAAGATGAGGAAGAAGAA---TTTGGTGAATTC 2926

RESULT 14
US-09-925-065A-133584
; Sequence 133584, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 133584
;   LENGTH: 575
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-133584

Alignment Scores:
Pred. No.:              4.79          Length:          575
Score:                  54.00         Matches:         10
Percent Similarity:     91.67%        Conservative:    1
Best Local Similarity:  83.33%        Mismatches:      1
Query Match:            44.26%        Indels:          0
DB:                     13            Gaps:            0

US-10-691-590-3 (1-26) x US-09-925-065A-133584 (1-575)

Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGly 14
              |||:::|||||||||   ||||||||||||||||||
Db         88 GAGGAAGAAGAGGAAGAGGAGGAGGAGACATTTGGT 123

RESULT 15
US-09-925-065A-854915
; Sequence 854915, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
```

; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 854915
; LENGTH: 625
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-854915

Alignment Scores:

Pred. No.:	5.24	Length:	625
Score:	54.00	Matches:	11
Percent Similarity:	61.54%	Conservative:	5
Best Local Similarity:	42.31%	Mismatches:	10
Query Match:	44.26%	Indels:	0
DB:	13	Gaps:	0

US-10-691-590-3 (1-26) x US-09-925-065A-854915 (1-625)

```
Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||:::   ||||||   :::||||||||||      |||            :::
Db        466 CGCAGGCAAGCAGAGGAGGCCGCAGAGGAAGAGGAGGGAGGCCCAGCAGAGGCCCTTCCC 525

Qy         21 AlaProLeu***ProGly 26
              :::|||:::   ||||||
Db        526 AGCCCTATCCTCCCCGGA 543
```

Search completed: November 5, 2005, 06:21:05
Job time : 278.165 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 5, 2005, 02:34:02 ; Search time 56.2957 Seconds
(without alignments)
755.710 Million cell updates/sec

Title: US-10-691-590-3
Perfect score: 122
Sequence: 1 RREQEEESEEETFGEFXQVXAPLXPG 26

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10691590/runat_02112005_120023_13203/app_query.fasta_1.462
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10691590_@CGN_1_1_93_@runat_02112005_120023_13203 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	117	95.9	1867	1	US-07-955-905A-1	Sequence 1, Appli
2	53	43.4	4756	4	US-09-949-016-4455	Sequence 4455, Ap
3	51	41.8	53737	4	US-09-949-016-16197	Sequence 16197, A
4	50.5	41.4	601	4	US-09-949-016-92026	Sequence 92026, A
c 5	50.5	41.4	254405	4	US-09-949-016-14381	Sequence 14381, A
6	50	41.0	1816	4	US-09-762-258-1	Sequence 1, Appli
7	49	40.2	601	4	US-09-949-016-65666	Sequence 65666, A
8	49	40.2	2677	4	US-09-949-016-1924	Sequence 1924, Ap
c 9	49	40.2	4474	3	US-09-221-017B-114	Sequence 114, App
10	49	40.2	19237	4	US-09-949-016-13666	Sequence 13666, A
c 11	49	40.2	19237	4	US-09-949-016-13666	Sequence 13666, A
12	48	39.3	330	4	US-09-489-039A-4753	Sequence 4753, Ap
c 13	48	39.3	420	4	US-09-489-039A-4835	Sequence 4835, Ap
c 14	48	39.3	790	3	US-08-998-416-421	Sequence 421, App
15	48	39.3	1548	4	US-09-543-681A-3556	Sequence 3556, Ap
c 16	48	39.3	1741	3	US-08-795-430-12	Sequence 12, Appl
c 17	48	39.3	1741	3	US-09-355-700-12	Sequence 12, Appl
c 18	48	39.3	1741	4	US-09-534-376A-12	Sequence 12, Appl
19	48	39.3	9051	4	US-09-949-016-12112	Sequence 12112, A
20	48	39.3	9052	4	US-09-949-016-15529	Sequence 15529, A
c 21	48	39.3	94133	4	US-09-949-016-11901	Sequence 11901, A
c 22	48	39.3	94133	4	US-09-949-016-12713	Sequence 12713, A
c 23	48	39.3	94135	4	US-09-949-016-15934	Sequence 15934, A
c 24	48	39.3	94135	4	US-09-949-016-15935	Sequence 15935, A
c 25	48	39.3	94135	4	US-09-949-016-15936	Sequence 15936, A
c 26	48	39.3	94135	4	US-09-949-016-15937	Sequence 15937, A
27	48	39.3	109250	4	US-09-949-016-12530	Sequence 12530, A
28	48	39.3	109251	4	US-09-949-016-17321	Sequence 17321, A
29	48	39.3	237510	4	US-09-949-016-14273	Sequence 14273, A
30	47.5	38.9	11558	4	US-09-949-016-14317	Sequence 14317, A
c 31	47	38.5	56	4	US-09-981-803-24	Sequence 24, Appl
c 32	47	38.5	57	3	US-08-894-511-18	Sequence 18, Appl
c 33	47	38.5	57	4	US-09-655-728-18	Sequence 18, Appl
34	47	38.5	264	4	US-09-248-796A-6960	Sequence 6960, Ap
35	47	38.5	309	4	US-09-513-999C-20344	Sequence 20344, A
c 36	47	38.5	601	4	US-09-949-016-61851	Sequence 61851, A
c 37	47	38.5	601	4	US-09-949-016-112656	Sequence 112656,
38	47	38.5	601	4	US-09-949-016-115267	Sequence 115267,
39	47	38.5	601	4	US-09-949-016-144537	Sequence 144537,
40	47	38.5	1053	4	US-09-248-796A-4335	Sequence 4335, Ap
c 41	47	38.5	1449	4	US-09-489-039A-2198	Sequence 2198, Ap
42	47	38.5	3937	4	US-10-164-595-7	Sequence 7, Appli
43	47	38.5	8355	3	US-08-406-030A-23	Sequence 23, Appl
c 44	47	38.5	9382	4	US-09-949-016-15306	Sequence 15306, A
c 45	47	38.5	44166	4	US-09-949-016-15829	Sequence 15829, A

ALIGNMENTS

RESULT 1
US-07-955-905A-1
; Sequence 1, Application US/07955905A
; Patent No. 5770433
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: RECOMBINANT 47 AND 31 kD COCOA PROTEINS AND
; TITLE OF INVENTION: PRECURSOR
; NUMBER OF SEQUENCES: 28
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/955,905A
; FILING DATE: 21-JAN-1993
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1867 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Theobroma cacao
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14..1711
US-07-955-905A-1

Alignment Scores:
Pred. No.: 1.4e-10 Length: 1867
Score: 117.00 Matches: 23
Percent Similarity: 88.46% Conservative: 0

```
Best Local Similarity:  88.46%          Mismatches:     3
Query Match:            95.90%          Indels:         0
DB:                     1               Gaps:           0

US-10-691-590-3 (1-26) x US-07-955-905A-1 (1-1867)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db       1322 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 1381

Qy         21 AlaProLeu***ProGly 26
              ||||||||||   ||||||
Db       1382 GCCCCATTGTCACCTGGT 1399

RESULT 2
US-09-949-016-4455
; Sequence 4455, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4455
;   LENGTH: 4756
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-4455

Alignment Scores:
Pred. No.:              34.9            Length:         4756
Score:                  53.00           Matches:        10
Percent Similarity:     76.47%          Conservative:   3
Best Local Similarity:  58.82%          Mismatches:     4
Query Match:            43.44%          Indels:         0
DB:                     4               Gaps:           0

US-10-691-590-3 (1-26) x US-09-949-016-4455 (1-4756)

Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal 19
              ::::::|||||||||   |||||||||   ||||||:::|||      |||
Db       3046 GACAAGGAAGAGGAAGACGAGGAGGAGGAATTTGGAGACTTTCGGCTTGTT 3096

RESULT 3
US-09-949-016-16197
; Sequence 16197, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16197
;   LENGTH: 53737
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-16197

Alignment Scores:
Pred. No.:              1.27e+03        Length:         53737
Score:                  51.00           Matches:        10
Percent Similarity:     75.00%          Conservative:   2
Best Local Similarity:  62.50%          Mismatches:     4
Query Match:            41.80%          Indels:         0
DB:                     4               Gaps:           0

US-10-691-590-3 (1-26) x US-09-949-016-16197 (1-53737)

Qy          4 GlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal 19
              :::|||||||||   |||||||||   ||||||:::|||      |||
Db      28422 AGGGAAGAGGAAGACGAGGAGGAGGAATTTGGAGACTTTCGGCTTGTT 28469

RESULT 4
US-09-949-016-92026
; Sequence 92026, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 92026
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-92026

Alignment Scores:
Pred. No.:              8.52            Length:         601
Score:                  50.50           Matches:        11
Percent Similarity:     53.12%          Conservative:   6
Best Local Similarity:  34.38%          Mismatches:     6
Query Match:            41.39%          Indels:         9
DB:                     4               Gaps:           1

US-10-691-590-3 (1-26) x US-09-949-016-92026 (1-601)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGlu---------------------------- 11
              :::|||||::::::   |||:::|||   |||
Db        178 CAGAGAGAGAGACAGGGAAAATGAATATGAACAGAAGAGATGAATGGAAAACTGAAAA 237

Qy         12 ThrPheGlyGluPhe***GlnVal***AlaProLeu 23
              |||||||||   |||   |||:::      |||:::
Db        238 ACTTTTGGAAATTTCACCCAAATTACTTACCCAGTT 273

RESULT 5
US-09-949-016-14381/c
; Sequence 14381, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
```

```
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14381
;   LENGTH: 254405
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-14381

Alignment Scores:
Pred. No.:              9.33e+03        Length:         254405
Score:                  50.50           Matches:        11
Percent Similarity:     53.12%          Conservative:   6
Best Local Similarity:  34.38%          Mismatches:     6
Query Match:            41.39%          Indels:         9
DB:                     4               Gaps:           1

US-10-691-590-3 (1-26) x US-09-949-016-14381 (1-254405)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGlu---------------------------- 11
              :::||||||::::::   |||:::|||   |||
Db      93186 CAGAGAGAGAGAGACAGGGAAAATGAATATGAACAGAAGAGATGAATGGAAAACTGAAAA 93127

Qy         12 ThrPheGlyGluPhe***GlnVal***AlaProLeu 23
              |||||||||   |||   |||:::      |||:::
Db      93126 ACTTTTGGAAATTTCACCCAAATTACTTACCCAGTT 93091

RESULT 6
US-09-762-258-1
; Sequence 1, Application US/09762258
; Patent No. 6830909
; GENERAL INFORMATION:
;  APPLICANT: Gout, Ivan
;  APPLICANT:  Hara, Kenta
;  APPLICANT:  Waterfield, Michael
;  APPLICANT:  Yonezawa, Kazu
;  APPLICANT:  Ludwig Institute for Cancer Research
;  TITLE OF INVENTION: Identification and Functional Characterization of a
;  TITLE OF INVENTION:  No. 6830909el Ribosomal S6 Protein Kinase
;  FILE REFERENCE: 40750-5002-US
;  CURRENT APPLICATION NUMBER: US/09/762,258
;  CURRENT FILING DATE:  2001-02-05
;  PRIOR APPLICATION NUMBER: PCT/US99/17595
;  PRIOR FILING DATE: 1999-08-04
;  PRIOR APPLICATION NUMBER: 60/095,268
;  PRIOR FILING DATE: 1998-08-04
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 1816
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (77)..(1561)
;   OTHER INFORMATION: p70(beta) S6 Kinase gene
US-09-762-258-1

Alignment Scores:
Pred. No.:              37.3            Length:         1816
Score:                  50.00           Matches:        10
Percent Similarity:     54.17%          Conservative:   3
Best Local Similarity:  41.67%          Mismatches:     11
Query Match:            40.98%          Indels:         0
DB:                     4               Gaps:           0

US-10-691-590-3 (1-26) x US-09-762-258-1 (1-1816)

Qy          2 ArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal***Ala 21
              ||||||:::|||   |||   |||   ||||||      |||:::   :::
Db         36 AGAGAGAGAGAGAGAGAGAGAGAGAGACTCGTGCCGAATGGCACGAGGCCGACGGG 95

Qy         22 ProLeu***Pro 25
              |||      |||
Db         96 CCCGCGGGGCCG 107

RESULT 7
US-09-949-016-65666
; Sequence 65666, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 65666
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-65666

Alignment Scores:
Pred. No.:              15.4            Length:         601
Score:                  49.00           Matches:        10
Percent Similarity:     80.00%          Conservative:   2
Best Local Similarity:  66.67%          Mismatches:     3
Query Match:            40.16%          Indels:         0
DB:                     4               Gaps:           0

US-10-691-590-3 (1-26) x US-09-949-016-65666 (1-601)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGlu 15
              |||||||||:::|||||||||:::   ||||||      ||||||
Db        498 CGGAGGGAGGAGGAAGAGGAAGCCGGCGAGGAGGCGGAGGGGGAG 542

RESULT 8
US-09-949-016-1924
; Sequence 1924, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1924
;   LENGTH: 2677
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-1924

Alignment Scores:
```

```
Pred. No.:              86.6          Length:        2677
Score:                  49.00         Matches:       10
Percent Similarity:     80.00%        Conservative:  2
Best Local Similarity:  66.67%        Mismatches:    3
Query Match:            40.16%        Indels:        0
DB:                     4             Gaps:          0

US-10-691-590-3 (1-26) x US-09-949-016-1924 (1-2677)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGlu 15
              |||||||||:::|||||||||:::   ||||||      ||||||
Db        199 CGGAGGGAGGAGGAAGAGGAAGCCGGCGAGGAGGCGGAGGGGGAG 243

RESULT 9
US-09-221-017B-114/c
; Sequence 114, Application US/09221017B
; Patent No. 6444799
;  GENERAL INFORMATION:
;    APPLICANT:  Ross, Bruce C.
;    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;    NUMBER OF SEQUENCES:  1120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP2911
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  114:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4474 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORYPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1...4474
US-09-221-017B-114

Alignment Scores:
Pred. No.:              157           Length:        4474
Score:                  49.00         Matches:       9
Percent Similarity:     78.57%        Conservative:  2
Best Local Similarity:  64.29%        Mismatches:    3
Query Match:            40.16%        Indels:        0
DB:                     3             Gaps:          0

US-10-691-590-3 (1-26) x US-09-221-017B-114 (1-4474)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGly 14
              |||   |||::::::||||||   |||||||||   ||||||
Db        579 AGAGAAGAGAGGAAAGAAGAAAAAGAAGAAGAGGTTTTGGA 538

RESULT 10
US-09-949-016-13666
; Sequence 13666, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13666
;   LENGTH: 19237
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-13666

Alignment Scores:
Pred. No.:              848           Length:        19237
Score:                  49.00         Matches:       10
Percent Similarity:     80.00%        Conservative:  2
Best Local Similarity:  66.67%        Mismatches:    3
Query Match:            40.16%        Indels:        0
DB:                     4             Gaps:          0

US-10-691-590-3 (1-26) x US-09-949-016-13666 (1-19237)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGlu 15
              |||||||||:::|||||||||:::   ||||||      ||||||
Db       2199 CGGAGGGAGGAGGAAGAGGAAGCCGGCGAGGAGGCGGAGGGGGAG 2243

RESULT 11
US-09-949-016-13666/c
; Sequence 13666, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13666
```

```
;   LENGTH: 19237
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-13666

Alignment Scores:
Pred. No.:                848             Length:         19237
Score:                    49.00           Matches:        10
Percent Similarity:       80.00%          Conservative:   2
Best Local Similarity:    66.67%          Mismatches:     3
Query Match:              40.16%          Indels:         0
DB:                       4               Gaps:           0

US-10-691-590-3 (1-26) x US-09-949-016-13666 (1-19237)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGlu 15
              ||||||||||:::|||||||||:::   ||||||      ||||||
Db      17293 CGGAGGGAGGAGGAAGAGGAAGCCGGCGAGGAGGCGGAGGGGGAG 17249

RESULT 12
US-09-489-039A-4753
; Sequence 4753, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 4753
;   LENGTH: 330
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-4753

Alignment Scores:
Pred. No.:                11.4            Length:         330
Score:                    48.00           Matches:        9
Percent Similarity:       54.17%          Conservative:   4
Best Local Similarity:    37.50%          Mismatches:     11
Query Match:              39.34%          Indels:         0
DB:                       4               Gaps:           0

US-10-691-590-3 (1-26) x US-09-489-039A-4753 (1-330)

Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal***AlaPro 22
              ||||||      :::   |||       |||   |||:::      ||||||   :::
Db        220 GAACAGCATGCGCAACAGGAGAATATGACTCTCGGGCAGCTGCGGCAGGTCATCAGCGAC 279

Qy         23 Leu***ProGly 26
              :::   ||||||
Db        280 ATCTATCCTGGC 291

RESULT 13
US-09-489-039A-4835/c
; Sequence 4835, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 4835
;   LENGTH: 420
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-4835

Alignment Scores:
Pred. No.:                15              Length:         420
Score:                    48.00           Matches:        9
Percent Similarity:       54.17%          Conservative:   4
Best Local Similarity:    37.50%          Mismatches:     11
Query Match:              39.34%          Indels:         0
DB:                       4               Gaps:           0

US-10-691-590-3 (1-26) x US-09-489-039A-4835 (1-420)

Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal***AlaPro 22
              ||||||      :::   |||       |||   |||:::      ||||||   :::
Db        192 GAACAGCATGCGCAACAGGAGAATATGACTCTCGGGCAGCTGCGGCAGGTCATCAGCGAC 133

Qy         23 Leu***ProGly 26
              :::   ||||||
Db        132 ATCTATCCTGGC 121

RESULT 14
US-08-998-416-421/c
; Sequence 421, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  421:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  790 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1311UP
US-08-998-416-421
```

Alignment Scores:

Pred. No.:	31.2	Length:	790
Score:	48.00	Matches:	10
Percent Similarity:	68.75%	Conservative:	1
Best Local Similarity:	62.50%	Mismatches:	5
Query Match:	39.34%	Indels:	0
DB:	3	Gaps:	0

US-10-691-590-3 (1-26) x US-08-998-416-421 (1-790)

```
Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***Gln 18
              |||:::|||||||||   |||||||||      ||||||      |||
Db        149 GAGGAAGAAGAGGAAGAAGAGGAAGAGGAAGAAGGAGAAGTATCACAG 102
```

RESULT 15
US-09-543-681A-3556
; Sequence 3556, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 3556
; LENGTH: 1548
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-3556

Alignment Scores:

Pred. No.:	68	Length:	1548
Score:	48.00	Matches:	9
Percent Similarity:	84.62%	Conservative:	2
Best Local Similarity:	69.23%	Mismatches:	2
Query Match:	39.34%	Indels:	0
DB:	4	Gaps:	0

US-10-691-590-3 (1-26) x US-09-543-681A-3556 (1-1548)

```
Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPhe 13
              ::::::|||||||||||||||   |||||||||   |||
Db        625 CAGAAAGAGCAAGAAGAAGAGGAAGAAGAGGAGGATTTC 663
```

Search completed: November 5, 2005, 06:01:44
Job time : 86.2957 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 5, 2005, 00:20:21 ; Search time 166.174 Seconds
(without alignments)
926.218 Million cell updates/sec

Title: US-10-691-590-3
Perfect score: 122
Sequence: 1 RREQEEESEEETFGEFXQVXAPLXPG 26

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10691590/runat_02112005_120021_13134/app_query.fasta_1.462
-DB=N_Geneseq_16Dec04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10691590_@CGN_1_1_586_@runat_02112005_120021_13134 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	117	95.9	1867	2	AAQ20377	Aaq20377 Sequence
c	2	65.5	53.7	6519	5	ABA16909	Aba16909 Human ner
c	3	59	48.4	1845	12	ACH91596	Ach91596 Human gen
	4	59	48.4	2365	12	ADQ18832	Adq18832 Human sof
	5	59	48.4	2365	13	ADQ38361	Adq38361 Human SNP
	6	59	48.4	2375	13	ADQ38360	Adq38360 Human SNP
	7	59	48.4	2382	12	ADQ23228	Adq23228 Human sof
c	8	59	48.4	4183	4	AAK74968	Aak74968 Human imm
c	9	59	48.4	4183	4	AAK75364	Aak75364 Human imm
	10	57	46.7	10266	2	AAT33007	Aat33007 Mouse SRY
c	11	57	46.7	205388	12	ADQ97560	Adq97560 Mouse can
c	12	55	45.1	2589	4	ABL09646	Abl09646 Drosophil
	13	54.5	44.7	4754	6	ABK52610	Abk52610 DNA encod
c	14	53	43.4	841	3	ABN81117	Abn81117 Shrimp po
	15	53	43.4	4414	10	ADC30694	Adc30694 Human nov
	16	53	43.4	4755	6	ABK52611	Abk52611 DNA encod
	17	53	43.4	4804	5	ABV30199	Abv30199 Human pro
c	18	53	43.4	110000	12	ADN46845_11	Continuation (12 o
	19	53	43.4	110000	12	ADN47591_08	Continuation (9 of
	20	53	43.4	110000	12	ADN47591_09	Continuation (10 o
c	21	53	43.4	110000	12	ADN46123_11	Continuation (12 o
	22	53	43.4	110000	12	ADN47209_08	Continuation (9 of
	23	53	43.4	110000	12	ADN47209_09	Continuation (10 o
c	24	53	43.4	110000	12	ADN46464_11	Continuation (12 o
	25	53	43.4	110000	12	ADN47960_08	Continuation (9 of
	26	53	43.4	110000	12	ADN47960_09	Continuation (10 o
	27	53	43.4	111206	11	ACN45152	Acn45152 Mouse gen
c	28	53	43.4	134841	11	ACN45172	Acn45172 Mouse gen
	29	52	42.6	403	4	ABA08483	Aba08483 Human L-p
	30	52	42.6	405	8	ABX62598	Abx62598 Arabidops
	31	52	42.6	1845	12	ADO62521	Ado62521 Transcrip
	32	52	42.6	3233	12	ADO62523	Ado62523 Transcrip
	33	52	42.6	147309	6	ABK49450	Abk49450 Human tra
	34	52	42.6	312477	12	ADP69744	Adp69744 Human ROC
	35	51.5	42.2	6953	12	ADO39645	Ado39645 Human lam
c	36	51.5	42.2	13572	9	AAD58743	Aad58743 Human tra
	37	51	41.8	939	5	AAS84137	Aas84137 DNA encod
c	38	51	41.8	5778	10	AAD55816	Aad55816 Micromono
c	39	51	41.8	17732	3	AAC68300	Aac68300 Lama2/APP
c	40	51	41.8	20623	3	AAC68294	Aac68294 Lama2/APP
c	41	51	41.8	22533	13	ADQ91713	Adq91713 Polyketid
c	42	51	41.8	31812	11	ACN44700	Acn44700 Mouse gen
	43	51	41.8	58837	6	ABK52612	Abk52612 Human Cla
c	44	51	41.8	60196	10	AAD55810	Aad55810 Micromono
	45	51	41.8	89736	13	ABD32966_10	Continuation (11 o

ALIGNMENTS

RESULT 1
AAQ20377
ID AAQ20377 standard; cDNA; 1867 BP.
XX
AC AAQ20377;
XX
DT 16-APR-1992 (first entry)
XX
DE Sequence of 67 kD T. cacao protein cDNA.
XX
KW Cocoa; flavour; vicilin; seed storage protein; ss.
XX
OS Theobroma cacao.
XX
FH Key Location/Qualifiers
FT CDS 14. .1714
FT /*tag= a
XX
PN WO9119801-A.
XX
PD 26-DEC-1991.
XX
PF 11-JUN-1990; 90GB-00013016.
XX
PR 11-JUN-1990; 90GB-00013016.
XX
PA (MRSC) MARS UK LTD.
XX
PI Spencer ME, Hodge R, Deakin EA, Ashton S;

XX
DR WPI; 1992-024418/03.
DR P-PSDB; AAR20181.
XX
PT Recombinant cocoa proteins - are responsible for flavour in cocoa beans
PT and produced in large quantities using yeast and bacterial expression
PT vectors.
XX
PS Claim 9; Fig 2; 59pp; English.
XX
CC The inventors claim a 67 kD and 31 kD T. cacao protein, and fragments,
CC and encoding DNAs. The 47 kD and 31 kD proteins are derived from the 67
CC kD precursor. T. cacao protein cDNA was detected in a cDNA library
CC prepared from immature cocoa beans RNA using a probe based on the AA
CC sequence of a CNBr peptide common to the 47 kD and 31 kD polypeptides.
CC Homology searches revealed close homologies between the 67 kD polypeptide
CC and the vicilins, which are seed storage proteins
XX
SQ Sequence 1867 BP; 578 A; 402 C; 460 G; 427 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	9.15e-10	Length:	1867
Score:	117.00	Matches:	23
Percent Similarity:	88.46%	Conservative:	0
Best Local Similarity:	88.46%	Mismatches:	3
Query Match:	95.90%	Indels:	0
DB:	2	Gaps:	0

US-10-691-590-3 (1-26) x AAQ20377 (1-1867)

```
Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db       1322 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 1381

Qy         21 AlaProLeu***ProGly 26
              |||||||||   ||||||
Db       1382 GCCCCATTGTCACCTGGT 1399
```

RESULT 2
ABA16909/c
ID ABA16909 standard; DNA; 6519 BP.
XX
AC ABA16909;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 9240.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001334.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.

```
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   20-OCT-2000; 2000US-0242221P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250391P.
PR   01-DEC-2000; 2000US-0251160P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system cancers
PT   and metastases.
XX
PS   Disclosure; SEQ ID NO 9240; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC   ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC   breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC   cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC   ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC   infectious diseases such as viral, bacterial, fungal and parasitic
CC   infections. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 6519 BP; 1661 A; 1577 C; 1542 G; 1739 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                1.66          Length:          6519
Score:                    65.50         Matches:         14
Percent Similarity:       59.26%        Conservative:    2
Best Local Similarity:    51.85%        Mismatches:      8
Query Match:              53.69%        Indels:          3
DB:                       5             Gaps:            1

US-10-691-590-3 (1-26) x ABA16909 (1-6519)

Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***Gln---------Val 19
              |||:::|||||||||   |||||||||   ||||||   |||   |||          :::
Db       3032 GAGGAGGAGGAAGAAGAGGAGGAAGAATTATTTGGCAATTTCTCACAAAGGTCAAACATG 2973

Qy         20 ***AlaProLeu***ProGly 26
                  |||       ||||||
Db       2972 GAGCCACCATTTGACCCAGGA 2952

RESULT 3
ACH91596/c
ID   ACH91596 standard; DNA; 1845 BP.
XX
AC   ACH91596;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Human genome derived single exon probe #24791.
XX
KW   Human; probe; ss; gene expression; single exon probe; microarray;
KW   alternative splicing event; genomic alteration.
XX
OS   Homo sapiens.
XX
PN   US2003194704-A1.
XX
PD   16-OCT-2003.
XX
PF   03-APR-2002; 2002US-00029386.
XX
PR   03-APR-2002; 2002US-00029386.
XX
PA   (PENN/) PENN S G.
PA   (RANK/) RANK D R.
PA   (HANZ/) HANZEL D K.
XX
PI   Penn SG,  Rank DR,  Hanzel DK;
XX
DR   WPI; 2004-119264/12.
XX
PT   New human genome-derived single exon nudleic acid probes useful for human
PT   gene expression analysis, for identifying or characterizing alternative
PT   splicing events, for assessing genomic alterations or as tools for
PT   surveying tissues.
XX
PS   Claim 1; SEQ ID NO 24791; 80pp; English.
XX
CC   The invention relates to a nucleic acid probe for measuring human gene
CC   expression, comprising any of the 27,400 fully defined nucleotide
CC   sequences in the specification, or their complements or fragments, and
```

```
CC   encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC   fully defined in the specification. The probe is a single exon probe that
CC   hybridises under high stringency conditions to a nucleic acid molecule
CC   expressed in human cells or tissues. Also included are a spatially-
CC   addressable set of single exon nucleic acid probes for measuring human
CC   gene expression (comprising a plurality of single exon nucleic acid
CC   probes cited above, where each of the plurality of probes is separately
CC   and addressably isolatable or amplifiable from the plurality), a single
CC   exon microarray for measuring human gene expression, a method of
CC   measuring human gene expression, a vector comprising the single exon
CC   probe cited above, an ORF-encoded peptide comprising at least 8
CC   contiguous amino acids of any of the above- mentioned amino acid
CC   sequences (optionally with conservative amino acid substitutions), an
CC   isolated antibody that binds specifically to a peptide cited above,
CC   methods of selling and/or licensing single exon probes or microarrays to
CC   a customer desiring to measure gene expression, a method of providing
CC   human gene expression data by subscription, and a computer-readable
CC   storage medium which contains a database having a plurality of records
CC   (each record including data on the expression of a single exon probe
CC   cited above. The probe, methods and apparatus are useful in gene
CC   expression analysis. The probes may be used as tools for surveying
CC   tissues to detect the presence of expressed messages that contain their
CC   specific exon, or in constructing genome-derived single exon microarrays.
CC   In addition, the probes are used in identifying and characterising
CC   alternative splicing events, in detecting and characterising gross
CC   alterations in the genomic locus that includes their exon, in assessing
CC   smaller genomic alterations, in priming the synthesis of nucleic acids,
CC   or in expressing the ORF-encoded peptide. The present sequence is a human
CC   single exon probe of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030194704
XX
SQ   Sequence 1845 BP; 258 A; 601 C; 442 G; 544 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              4.9          Length:        1845
Score:                  59.00        Matches:       12
Percent Similarity:     64.00%       Conservative:  4
Best Local Similarity:  48.00%       Mismatches:    9
Query Match:            48.36%       Indels:        0
DB:                     12           Gaps:          0

US-10-691-590-3 (1-26) x ACH91596 (1-1845)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              :::   |||:::|||||||||   :::||||||          |||      :::|||
Db        442 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 383

Qy         21 AlaProLeu***Pro 25
              |||||||||   |||
Db        382 GCCCCACTGAGCCCA 368

RESULT 4
ADQ18832
ID   ADQ18832 standard; DNA; 2365 BP.
XX
AC   ADQ18832;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Human soft tissue sarcoma-upregulated DNA - SEQ ID 1651.
XX
KW   soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO2004048938-A2.
XX
PD   10-JUN-2004.
XX
PF   26-NOV-2003; 2003WO-US038193.
XX
PR   26-NOV-2002; 2002US-0429739P.
XX
PA   (PROT-) PROTEIN DESIGN LABS INC.
XX
PI   Aziz N,  Ginsburg WM,  Zlotnik A;
XX
DR   WPI; 2004-441208/41.
XX
PT   Early detection of soft tissue sarcoma comprises determining expression
PT   of a gene in a first soft tissue sample and a normal soft tissue sample
PT   and comparing the gene expression, also useful in treating soft tissue
PT   sarcoma.
XX
PS   Example 2; SEQ ID NO 1651; 210pp; English.
XX
CC   The invention relates to a novel method for detecting soft tissue sarcoma
CC   which comprises obtaining a first soft tissue sample from an individual
CC   and a normal soft tissue sample from the same or different individual,
CC   determining the expression of a gene in both samples and comparing the
CC   expression of the gene in both soft tissue samples, where a higher level
CC   of protein expression in the first soft tissue sample indicates the
CC   presence of soft tissue sarcoma. The method of the invention has
CC   cytostatic applications and may be useful for detecting soft tissue
CC   sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC   acid sequences may be useful in diagnostic and screening applications.
CC   The current sequence is that of a human soft tissue sarcoma-upregulated
CC   DNA of the invention. The current sequence is not shown within the
CC   specification per se but was submitted in CD format by the inventor.
XX
SQ   Sequence 2365 BP; 682 A; 606 C; 731 G; 346 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              6.49         Length:        2365
Score:                  59.00        Matches:       12
Percent Similarity:     64.00%       Conservative:  4
Best Local Similarity:  48.00%       Mismatches:    9
Query Match:            48.36%       Indels:        0
DB:                     12           Gaps:          0

US-10-691-590-3 (1-26) x ADQ18832 (1-2365)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              :::   |||:::|||||||||   :::||||||          |||      :::|||
Db       1788 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 1847

Qy         21 AlaProLeu***Pro 25
              |||||||||   |||
Db       1848 GCCCCACTGAGCCCA 1862

RESULT 5
ADQ38361
ID   ADQ38361 standard; DNA; 2365 BP.
XX
AC   ADQ38361;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human SNP containing myocardial infarction-associated gene, SEQ ID 24.
XX
KW   Myocardial infarction; detection; single nucleotide polymorphism; SNP;
KW   cardiant; gene therapy; human; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004058052-A2.
XX
PD   15-JUL-2004.
XX
PF   22-DEC-2003; 2003WO-US040978.
XX
PR   20-DEC-2002; 2002US-0434778P.
PR   10-MAR-2003; 2003US-0453135P.
```

PR 30-APR-2003; 2003US-0466412P.
PR 23-SEP-2003; 2003US-0504955P.
XX
PA (APPL-) APPLERA CORP.
XX
PI Cargill M, Devlin JJ, Iakoubova O;
XX
DR WPI; 2004-533949/51.
DR P-PSDB; ADQ39189.
XX
PT Identifying an individual who has an altered risk for developing
PT myocardial infarction by detecting a single nucleotide polymorphism in
PT the individual's nucleic acids.
XX
PS Claim 7; SEQ ID NO 24; 145pp; English.
XX
CC The invention relates to a novel method for identifying an individual who
CC has an altered risk for developing myocardial infarction. The method
CC comprises detecting a single nucleotide polymorphism (SNP) in any one of
CC the nucleotide sequences given in the specification in the individual's
CC nucleic acids, where the presence of the SNP is correlated with an
CC altered risk for myocardial infarction in the individual. The invention
CC further comprises: an isolated nucleic acid molecule comprising at least
CC 8 contiguous nucleotides where one of the nucleotides is an SNP given in
CC the specification or its complement and encoding any one of the amino
CC acid sequences given in the specification; an isolated polypeptide
CC comprising an amino acid sequence given in the specification; an antibody
CC that specifically binds to the polypeptide or its antigen-binding
CC fragment; an amplified polynucleotide containing an SNP given in the
CC specification and which is between about 16 and 1000 nucleotides in
CC length; a kit for detecting an SNP in a nucleic acid, comprising the
CC polynucleotide, a buffer and an enzyme; a method of detecting an SNP in a
CC nucleic acid molecule; a method of detecting a variant polypeptide; and a
CC method for identifying an agent useful in treating or preventing
CC myocardial infarction. The novel detection method has cardiant activity.
CC The nucleic acids of the invention may be used in gene therapy. The
CC method is useful in identifying an individual who has an increased or
CC decreased risk for developing myocardial infarction and for preparing a
CC composition for treating or preventing myocardial infarction. This
CC polynucleotide sequence represents a human myocardial infarction-
CC associated gene containing one or more SNP's of the invention. Note: This
CC sequence was not shown in the specification. The sequence has come from
CC an electronic sequence listing downloaded from the WIPO website.
XX
SQ Sequence 2365 BP; 681 A; 604 C; 728 G; 344 T; 0 U; 8 Other;

Alignment Scores:

Pred. No.:	6.49	Length:	2365
Score:	59.00	Matches:	12
Percent Similarity:	64.00%	Conservative:	4
Best Local Similarity:	48.00%	Mismatches:	9
Query Match:	48.36%	Indels:	0
DB:	13	Gaps:	0

US-10-691-590-3 (1-26) x ADQ38361 (1-2365)

```
Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              :::  |||:::|||||||||   :::||||||        |||      :::|||
Db       1788 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 1847

Qy         21 AlaProLeu***Pro 25
              |||||||||   |||
Db       1848 GCCCCACTGAGCCCA 1862
```

RESULT 6
ADQ38360
ID ADQ38360 standard; DNA; 2375 BP.
XX
AC ADQ38360;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human SNP containing myocardial infarction-associated gene, SEQ ID 23.
XX
KW Myocardial infarction; detection; single nucleotide polymorphism; SNP;
KW cardiant; gene therapy; human; gene; ds.
XX
OS Homo sapiens.
XX
PN WO2004058052-A2.
XX
PD 15-JUL-2004.
XX
PF 22-DEC-2003; 2003WO-US040978.
XX
PR 20-DEC-2002; 2002US-0434778P.
PR 10-MAR-2003; 2003US-0453135P.
PR 30-APR-2003; 2003US-0466412P.
PR 23-SEP-2003; 2003US-0504955P.
XX
PA (APPL-) APPLERA CORP.
XX
PI Cargill M, Devlin JJ, Iakoubova O;
XX
DR WPI; 2004-533949/51.
DR P-PSDB; ADQ39188.
XX
PT Identifying an individual who has an altered risk for developing
PT myocardial infarction by detecting a single nucleotide polymorphism in
PT the individual's nucleic acids.
XX
PS Claim 7; SEQ ID NO 23; 145pp; English.
XX
CC The invention relates to a novel method for identifying an individual who
CC has an altered risk for developing myocardial infarction. The method
CC comprises detecting a single nucleotide polymorphism (SNP) in any one of
CC the nucleotide sequences given in the specification in the individual's
CC nucleic acids, where the presence of the SNP is correlated with an
CC altered risk for myocardial infarction in the individual. The invention
CC further comprises: an isolated nucleic acid molecule comprising at least
CC 8 contiguous nucleotides where one of the nucleotides is an SNP given in
CC the specification or its complement and encoding any one of the amino
CC acid sequences given in the specification; an isolated polypeptide
CC comprising an amino acid sequence given in the specification; an antibody
CC that specifically binds to the polypeptide or its antigen-binding
CC fragment; an amplified polynucleotide containing an SNP given in the
CC specification and which is between about 16 and 1000 nucleotides in
CC length; a kit for detecting an SNP in a nucleic acid, comprising the
CC polynucleotide, a buffer and an enzyme; a method of detecting an SNP in a
CC nucleic acid molecule; a method of detecting a variant polypeptide; and a
CC method for identifying an agent useful in treating or preventing
CC myocardial infarction. The novel detection method has cardiant activity.
CC The nucleic acids of the invention may be used in gene therapy. The
CC method is useful in identifying an individual who has an increased or
CC decreased risk for developing myocardial infarction and for preparing a
CC composition for treating or preventing myocardial infarction. This
CC polynucleotide sequence represents a human myocardial infarction-
CC associated gene containing one or more SNP's of the invention. Note: This
CC sequence was not shown in the specification. The sequence has come from
CC an electronic sequence listing downloaded from the WIPO website.
XX
SQ Sequence 2375 BP; 685 A; 607 C; 729 G; 346 T; 0 U; 8 Other;

Alignment Scores:

Pred. No.:	6.52	Length:	2375
Score:	59.00	Matches:	12
Percent Similarity:	64.00%	Conservative:	4
Best Local Similarity:	48.00%	Mismatches:	9
Query Match:	48.36%	Indels:	0
DB:	13	Gaps:	0

US-10-691-590-3 (1-26) x ADQ38360 (1-2375)

```
Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              :::  |||:::|||||||||   :::||||||        |||      :::|||
Db       1788 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 1847
```

```
Qy         21 AlaProLeu***Pro 25
              |||||||||   |||
Db       1848 GCCCCACTGAGCCCA 1862

RESULT 7
ADQ23228
ID   ADQ23228 standard; DNA; 2382 BP.
XX
AC   ADQ23228;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Human soft tissue sarcoma-upregulated DNA - SEQ ID 6048.
XX
KW   soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO2004048938-A2.
XX
PD   10-JUN-2004.
XX
PF   26-NOV-2003; 2003WO-US038193.
XX
PR   26-NOV-2002; 2002US-0429739P.
XX
PA   (PROT-) PROTEIN DESIGN LABS INC.
XX
PI   Aziz N,  Ginsburg WM,  Zlotnik A;
XX
DR   WPI; 2004-441208/41.
XX
PT   Early detection of soft tissue sarcoma comprises determining expression
PT   of a gene in a first soft tissue sample and a normal soft tissue sample
PT   and comparing the gene expression, also useful in treating soft tissue
PT   sarcoma.
XX
PS   Example 2; SEQ ID NO 6048; 210pp; English.
XX
CC   The invention relates to a novel method for detecting soft tissue sarcoma
CC   which comprises obtaining a first soft tissue sample from an individual
CC   and a normal soft tissue sample from the same or different individual,
CC   determining the expression of a gene in both samples and comparing the
CC   expression of the gene in both soft tissue samples, where a higher level
CC   of protein expression in the first soft tissue sample indicates the
CC   presence of soft tissue sarcoma. The method of the invention has
CC   cytostatic applications and may be useful for detecting soft tissue
CC   sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC   acid sequences may be useful in diagnostic and screening applications.
CC   The current sequence is that of a human soft tissue sarcoma-upregulated
CC   DNA of the invention. The current sequence is not shown within the
CC   specification per se but was submitted in CD format by the inventor.
XX
SQ   Sequence 2382 BP; 668 A; 594 C; 708 G; 323 T; 0 U; 89 Other;

Alignment Scores:
Pred. No.:              6.55          Length:          2382
Score:                  59.00         Matches:         12
Percent Similarity:     64.00%        Conservative:    4
Best Local Similarity:  48.00%        Mismatches:      9
Query Match:            48.36%        Indels:          0
DB:                     12            Gaps:            0

US-10-691-590-3 (1-26) x ADQ23228 (1-2382)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              :::   |||:::|||||||||   :::||||||      |||     :::|||
Db       1788 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 1847

Qy         21 AlaProLeu***Pro 25
              |||||||||   |||

Db       1848 GCCCCACTGAGCCCA 1862

RESULT 8
AAK74968/c
ID   AAK74968 standard; DNA; 4183 BP.
XX
AC   AAK74968;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:29780.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
```

PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 29780; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 4183 BP; 659 A; 1349 C; 1068 G; 1107 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	12.4	Length:	4183
Score:	59.00	Matches:	12
Percent Similarity:	64.00%	Conservative:	4
Best Local Similarity:	48.00%	Mismatches:	9
Query Match:	48.36%	Indels:	0
DB:	4	Gaps:	0

US-10-691-590-3 (1-26) x AAK74968 (1-4183)

```
Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ::   |||:::|||||||||   :::||||||         |||      :::|||
Db       2422 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 2363

Qy         21 AlaProLeu***Pro 25
              |||||||||   |||
Db       2362 GCCCCACTGAGCCCA 2348
```

RESULT 9
AAK75364/c
ID AAK75364 standard; DNA; 4183 BP.
XX
AC AAK75364;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:30176.
XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.

```
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 30176; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 4183 BP; 659 A; 1349 C; 1068 G; 1107 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              12.4          Length:         4183
Score:                  59.00         Matches:        12
Percent Similarity:     64.00%        Conservative:   4
Best Local Similarity:  48.00%        Mismatches:     9
Query Match:            48.36%        Indels:         0
DB:                     4             Gaps:           0

US-10-691-590-3 (1-26) x AAK75364 (1-4183)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              :::   |||:::|||||||||   :::||||||      |||     :::|||
Db       2422 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 2363

Qy         21 AlaProLeu***Pro 25
              |||||||||   |||
Db       2362 GCCCCACTGAGCCCA 2348

RESULT 10
AAT33007
ID   AAT33007 standard; DNA; 10266 BP.
XX
AC   AAT33007;
XX
DT   23-OCT-1996  (first entry)
XX
DE   Mouse SRY-related gene.
XX
KW   Mouse; SRY; primer; PCR; polymerase chain reaction; amplification; probe;
KW   HMG box; human; bovine; sex; animal; birth; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             7148. .8335
FT                   /*tag=  a
FT                   /product= "SRY-related protein"
XX
PN   JP08154685-A.
XX
PD   18-JUN-1996.
XX
PF   30-NOV-1994;   94JP-00319525.
XX
PR   30-NOV-1994;   94JP-00319525.
XX
PA   (KACH-) KACHIKU JUSEIRAN ISHOKU GIKUTSU KENKYUKU.
XX
DR   WPI; 1996-336575/34.
DR   P-PSDB; AAW03474.
XX
PT   Bovine and mouse Sry-related DNA - useful for detecting e.g. the sex of
PT   unborn animals.
XX
PS   Claim 1; Page 10-14; 21pp; Japanese.
XX
CC   This is the nucleotide sequence of a gene encoding a mouse SRY-related
CC   protein. The gene was isolated from a mouse genomic library using a cDNA
CC   fragment amplified by primers AAT33009-10 as a probe. The screen isolated
CC   4 EcoRI fragments of 2.3, 2.8, 3.5 and 1.5 kb covering the gene. Sequence
CC   analysis revealed a 240 bp HMG box sequence between bases 7154-7393.
CC   Similarity with the human SRY HMG box sequence resulted in primers being
CC   generated to amplify the human SRY HMG box sequence for use as a probe to
CC   isolate the bovine SRY-related gene (AAT33008). The mouse and bovine
CC   genes are useful for determining the sex of an animal prior to birth
XX
SQ   Sequence 10266 BP; 2958 A; 2202 C; 2228 G; 2878 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              74.4          Length:         10266
Score:                  57.00         Matches:        12
Percent Similarity:     78.95%        Conservative:   3
Best Local Similarity:  63.16%        Mismatches:     4
Query Match:            46.72%        Indels:         0
DB:                     2             Gaps:           0

US-10-691-590-3 (1-26) x AAT33007 (1-10266)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal 19
              ||||||::::::|||||||||   ||||||||||||   |||:::      ||||||
Db       4724 AGGAGGAAGAAAGAGGAGGAGGAGGAGGAGGAGACAGAGGGAGATATGTAGCAAGTA 4780

RESULT 11
ADQ97560/c
ID   ADQ97560 standard; DNA; 205388 BP.
XX
AC   ADQ97560;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Mouse cancer associated sequence MD09-014, SEQ ID 537.
XX
KW   Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.
XX
OS   Mus musculus.
XX
PN   WO2004060304-A2.
XX
PD   22-JUL-2004.
XX
PF   22-DEC-2003; 2003WO-US041389.
XX
PR   27-DEC-2002; 2002US-00330773.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-543781/52.
XX
PT   New isolated cancer associated nucleic acids comprising at least 10
PT   contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT   cancers such as leukemia and lymphoma.
```

```
XX
PS   Claim 1; SEQ ID NO 537; 199pp; English.
XX
CC   The present invention relates to cancer associated sequences (ADQ97025-
CC   ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC   treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic formate directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 205388 BP; 55423 A; 41274 C; 44636 G; 63134 T; 0 U; 921 Other;

Alignment Scores:
Pred. No.:              2.23e+03        Length:         205388
Score:                  57.00           Matches:        13
Percent Similarity:     68.00%          Conservative:   4
Best Local Similarity:  52.00%          Mismatches:     6
Query Match:            46.72%          Indels:         2
DB:                     12              Gaps:           1

US-10-691-590-3 (1-26) x ADQ97560 (1-205388)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ::::::|||:::|||||||||   |||||||||      ||||||      :::|||
Db      25178 AAGAAGGAGAAGGAGGAGGAAGAAGAGGAGGAG------GGGGAAGAGGAGGAGGTCCTC 25125

Qy         21 AlaProLeu***Pro 25
                 ||||||   |||
Db      25124 CCCCCTCTCCTCCCC 25110

RESULT 12
ABL09646/c
ID   ABL09646 standard; cDNA; 2589 BP.
XX
AC   ABL09646;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 23420.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB65543.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 23420; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 2589 BP; 799 A; 533 C; 616 G; 641 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              33.8            Length:         2589
Score:                  55.00           Matches:        12
Percent Similarity:     57.69%          Conservative:   3
Best Local Similarity:  46.15%          Mismatches:     7
Query Match:            45.08%          Indels:         4
DB:                     4               Gaps:           1

US-10-691-590-3 (1-26) x ABL09646 (1-2589)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              |||:::|||:::|||   |||   |||   |||                   |||:::
Db        578 CGCAAAGAGAGAGAGAGAGAGAGAGAAAGCGAACGGATG-----------CAAATTGCT 531

Qy         21 AlaProLeu***ProGly 26
                 |||||||||   ||||||
Db        530 GCTCCTTTGGCTCCGGGT 513

RESULT 13
ABK52610
ID   ABK52610 standard; DNA; 4754 BP.
XX
AC   ABK52610;
XX
DT   27-AUG-2002  (first entry)
XX
DE   DNA encoding Xenopus Claspin protein.
XX
KW   ChK1 protein; SQ/TQ motif; isoelectric point; cell cycle progression;
KW   nuclear localisation signal; DNA replication checkpoint; benign neoplasm;
KW   cell proliferative disorder; malignant neoplasm; frog; claspin; gene; ds.
XX
OS   Xenopus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             71. .3928
FT                   /*tag=  a
FT                   /product= "Xenopus Claspin protein"
XX
PN   WO200233115-A2.
XX
PD   25-APR-2002.
XX
PF   17-OCT-2001; 2001WO-US032316.
XX
PR   17-OCT-2000; 2000US-0241246P.
XX
PA   (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
PI   Kumagai A,  Dunphy WG;
XX
DR   WPI; 2002-454610/48.
DR   P-PSDB; AAU97586.
XX
PT   Novel Claspin polypeptide specifically interacting with chK1 protein
PT   useful for identifying compound that modulates cell cycle progression and
PT   for treating cell proliferative disorder like neoplasm.
XX
PS   Claim 7; Fig 1; 97pp; English.
XX
CC   The present invention relates to a new substantially pure Claspin
CC   polypeptide that specifically interacts with a ChK1 protein, having SQ/TQ
CC   motifs, an isoelectric point of 4.5 and at least one nuclear localisation
CC   signal. The method of the invention is useful for identifying a compound
CC   that modulates cell cycle progression and for modulating cell cycle
CC   progression in a cell. The invention is useful for the proper operation
```

CC of DNA replication checkpoint in the cell cycle. The method is also
CC useful for treating a disorder associated with cell cycle progression
CC e.g. cell proliferative disorder such as benign or malignant neoplasm.
CC The molecules of the invention are also useful for detecting the altered
CC levels of claspin expression. The present nucleic acid sequence encodes
CC the Xenopus Claspin protein of the invention
XX
SQ Sequence 4754 BP; 1658 A; 832 C; 1066 G; 1198 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	81.6	Length:	4754
Score:	54.50	Matches:	12
Percent Similarity:	86.67%	Conservative:	1
Best Local Similarity:	80.00%	Mismatches:	1
Query Match:	44.67%	Indels:	1
DB:	6	Gaps:	1

US-10-691-590-3 (1-26) x ABK52610 (1-4754)

```
Qy          2 ArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe 16
              ||||||||||||||:::   |||||||||   ||||||||||||
Db       2885 AGAGAACAGGAAGAAGATGAGGAAGAAGAA---TTTGGTGAATTC 2926
```

RESULT 14
ABN81117/c
ID ABN81117 standard; DNA; 841 BP.
XX
AC ABN81117;
XX
DT 16-JUL-2002 (first entry)
XX
DE Shrimp polynucleotide SEQ ID NO 121.
XX
KW Giant black tiger prawn; Penaeus monodon; pacific white shrimp;
KW Litopenaeus vannamei; shrimp; microsatellite sequence; genome mapping;
KW Taura Syndrome Virus; TSV; infection; ds.
XX
OS Litopenaeus vannamei.
XX
PN WO200034476-A2.
XX
PD 15-JUN-2000.
XX
PF 10-DEC-1999; 99WO-US029571.
XX
PR 10-DEC-1998; 98US-0111670P.
XX
PA (TUFT) TUFTS COLLEGE.
XX
PI Alcivar-Warren A, Xu Z, Dhar AK, Fan Y, Meehan D, Garcia DK;
XX
DR WPI; 2000-423422/36.
XX
PT Polynucleotides of shrimp are useful for identifying, mapping and
PT characterizing of the genome of various species of shrimp.
XX
PS Claim 1; Page 109; 120pp; English.
XX
CC The invention relates to an isolated polynucleotide (I) of the giant
CC black tiger prawn, Penaeus monodon or expressed sequence tags of the
CC pacific white shrimp, Litopenaeus vannemei (ABN80997-ABN81172), both
CC containing microsatellites sequences including those P. monodon
CC microsatellite sequences given in GenBank AF077550-AF077598. (I), the
CC complementary sequence or fragment and the encoded polypeptide are useful
CC for mapping of the genome of various species of shrimp. Mapping the
CC genome of Penaeus is useful for determining whether a test shrimp,
CC preferably Litopenaeus vannemei, has a genotype associated with a
CC phenotypic trait such as resistance to Taura Syndrome Virus (TSV)
CC infection
XX
SQ Sequence 841 BP; 98 A; 290 C; 78 G; 210 T; 0 U; 165 Other;

Alignment Scores:

Pred. No.:	20.4	Length:	841
Score:	53.00	Matches:	10
Percent Similarity:	80.00%	Conservative:	2
Best Local Similarity:	66.67%	Mismatches:	3
Query Match:	43.44%	Indels:	0
DB:	3	Gaps:	0

US-10-691-590-3 (1-26) x ABN81117 (1-841)

```
Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGlu 15
              |||  |||:::|||||||||  ||||||||||:::  ||||||
Db        239 AGGGAGGAGGAGGAGGAAGAGAAAGAAGAGGAAAGCCATGGGGAA 195
```

RESULT 15
ADC30694
ID ADC30694 standard; cDNA; 4414 BP.
XX
AC ADC30694;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human novel cDNA sequence, SEQ ID NO:776.
XX
KW Human; diagnostic; drug screening; forensics; gene mapping;
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
KW ulcers; osteoporosis; autoimmune disease; cancer;
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;
KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
KW gene therapy; chromosome 1; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003029271-A2.
XX
PD 10-APR-2003.
XX
PF 24-SEP-2002; 2002WO-US030474.
XX
PR 24-SEP-2001; 2001US-0324631P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
PI Haley-Vicente D, Drmanac RT;
XX
DR WPI; 2003-371981/35.
DR P-PSDB; ADC31665.
XX
PT New polynucleotide and polypeptide useful for diagnosing, preventing or
PT treating conditions such as neurodegenerative diseases, anemias, platelet
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
PT cancer.
XX
PS Claim 1; SEQ ID NO 776; 1185pp; English.
XX
CC The invention relates to 971 novel human cDNA sequences (ADC29919-
CC ADC30889)and the polypeptides they encode (ADC30890-ADC31860). The
CC invention also relates to nucleic acid sequences over 99% identical with
CC the novel human cDNAs. The invention additionally encompasses expression
CC vectors and host cells comprising a nucleic acid of the invention; the
CC recombinant production of a polypeptide of the invention; an antibody
CC against a polypeptide of the invention; a method of detecting
CC polynucleotides or polypeptides of the invention; and methods of
CC identifying a compound which binds to a polypeptide of the invention. The
CC invention further discloses methods of peventing, treating or
CC ameliorating a medical condition; kits comprising polynucleotide probes
CC and/or monoclonal antibodies for carrying out the methods of the
CC invention; methods for the identification of compounds that modulate the
CC expression or activity of the polynucleotide and/or polypeptide; and 767
CC contig sequences corresponding to the cDNA sequences of the invention

CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
CC -ADC33394). The nucleic acids and polypeptides of the invention are
CC useful in diagnostics, drug screening, forensics, gene mapping, in the
CC identification of mutations responsible for genetic disorders or other
CC traits, for assessing biodiversity, and in producing many other types of
CC data and products dependent on DNA and amino acid sequences. They are
CC also used for treating diseases such as Parkinson's disease, Alzheimer's
CC disease and other neurodegenerative diseases, anaemia, platelet
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primers, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a specifically
CC claimed human cDNA sequence of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4414 BP; 1497 A; 828 C; 1167 G; 922 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	134	Length:	4414
Score:	53.00	Matches:	10
Percent Similarity:	76.47%	Conservative:	3
Best Local Similarity:	58.82%	Mismatches:	4
Query Match:	43.44%	Indels:	0
DB:	10	Gaps:	0

US-10-691-590-3 (1-26) x ADC30694 (1-4414)

```
Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal 19
              ::::::|||||||||   |||||||||   ||||||:::|||      |||
Db       3236 GACAAGGAAGAGGAAGACGAGGAGGAGGAATTTGGAGACTTTCGGCTTGTT 3286
```

Search completed: November 5, 2005, 03:06:51
Job time : 193.174 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 5, 2005, 00:24:30 ; Search time 1280.78 Seconds
(without alignments)
983.645 Million cell updates/sec

Title: US-10-691-590-3
Perfect score: 122
Sequence: 1 RREQEEESEEETFGEFXQVXAPLXPG 26

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US10691590/runat_02112005_120021_13145/app_query.fasta_1.462
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10691590_@CGN_1_1_3851_@runat_02112005_120021_13145 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	117	95.9	1867	6	A20606	A20606 67kD protei
2	117	95.9	3318	8	TCCSV	X62625 T.cacao csv
3	82	67.2	1380	8	TCCSVSV	X62626 T.cacao csv
4	66	54.1	242889	2	AC095192	AC095192 Rattus no
c 5	66	54.1	252743	2	AC136578	AC136578 Rattus no
c 6	65.5	53.7	37492	9	AC010649	AC010649 Homo sapi
7	61	50.0	171432	2	AC102322	AC102322 Mus muscu
8	60	49.2	214464	2	AC121397	AC121397 Rattus no
9	60	49.2	236863	2	AC097079	AC097079 Rattus no
10	60	49.2	242913	9	AC115994	AC115994 Homo sapi
11	59	48.4	2332	9	BC069795	BC069795 Homo sapi
12	59	48.4	2332	9	BC069802	BC069802 Homo sapi
13	59	48.4	2365	6	CQ715312	CQ715312 Sequence
14	59	48.4	2365	9	HUMHCPB	M60052 Human histi
c 15	59	48.4	42003	2	AC145716	AC145716 Homo sapi
c 16	59	48.4	42920	2	AC145702	AC145702 Homo sapi
17	59	48.4	117628	9	AC008891	AC008891 Homo sapi
18	59	48.4	164635	2	AC143346	AC143346 Homo sapi
19	59	48.4	184579	2	AC105976	AC105976 Mus muscu
20	59	48.4	192387	10	AC151578	AC151578 Mus muscu
c 21	59	48.4	211763	10	AC123558	AC123558 Mus muscu
22	59	48.4	219784	10	AC112792	AC112792 Mus muscu
23	58	47.5	1828	9	AK129876	AK129876 Homo sapi
c 24	57	46.7	2490	8	MCPPC1A	X63774 M.crystalli
25	57	46.7	10266	6	E11536	E11536 DNA from Y
26	57	46.7	163378	8	AP004087	AP004087 Oryza sat
c 27	57	46.7	191261	2	AC074304	AC074304 Mus muscu
c 28	57	46.7	224138	10	AC109243	AC109243 Mus muscu
c 29	57	46.7	229813	2	AC117613	AC117613 Mus muscu
c 30	56.5	46.3	221983	2	AC122881	AC122881 Mus muscu
31	56	45.9	64318	2	AC100364	AC100364 Mus muscu
32	56	45.9	144555	2	AC037451	AC037451 Homo sapi
33	56	45.9	166601	2	AC132753	AC132753 Rattus no
c 34	56	45.9	175629	9	AC097173	AC097173 Homo sapi
c 35	56	45.9	178209	2	AC150629	AC150629 Papio anu
c 36	56	45.9	178601	9	HSJ1022P6	AL109935 Human DNA
37	56	45.9	183935	2	AC112962	AC112962 Mus muscu
c 38	56	45.9	192562	10	AC132373	AC132373 Mus muscu
39	56	45.9	199875	9	AC134407	AC134407 Homo sapi
c 40	56	45.9	206630	2	AC150604	AC150604 Callithri
c 41	56	45.9	207404	2	AC149189	AC149189 Papio anu
c 42	56	45.9	216322	2	AC151686	AC151686 Gallus ga
c 43	56	45.9	235706	2	AC132182	AC132182 Rattus no
44	56	45.9	247402	10	AC125187	AC125187 Mus muscu
c 45	55.5	45.5	110000	1	RME591985_02	Continuation (3 of

ALIGNMENTS

RESULT 1
A20606
LOCUS A20606 1867 bp DNA linear PAT 27-JUL-1994
DEFINITION 67kD protein.
ACCESSION A20606
VERSION A20606.1 GI:579221
KEYWORDS .
SOURCE Theobroma cacao (cacao)
ORGANISM Theobroma cacao
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae; Theobroma.
REFERENCE 1 (bases 1 to 1867)
AUTHORS .
TITLE RECOMBINANT 47 AND 31kD COCOA PROTEINS AND PRECURSOR
JOURNAL Patent: WO 9119801-A 6 26-DEC-1991;
FEATURES Location/Qualifiers
source 1. .1867
/organism="Theobroma cadao"
/mol_type="unassigned DNA"
/db_xref="taxon:3641"
gene 14. .1714
/gene="67kD protein"
CDS 14. .1714
/gene="67kD protein"
/codon_start=1
/protein_id="CAA01512.1"

```
                     /db_xref="GI:579222"
                     /translation="MVISKSPFIVLIFSLLLSFALLCSGVSAYGRKQYERDPRQQYEQ
                     CQRRCESEATEEREQEQCEQRCEREYKEQQRQQEEELQRQYQQCQGRCQEQQQGQREQ
                     QQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQRNNPYYFPKRRSFQTRFRDEE
                     GNFKILQRFAENSPPLKGINDYRLAMFEANPNTFILPHHCDAEAIYFVTNGKGTITFV
                     THENKESYNVQRGTVVSVPAGSTVYVVSQDNQEKLTIAVLALPVNSPGKYELFFPAGN
                     NKPESYYGAFSYEVLETVFNTQREKLEEILEEQRGQKRQQGQQGMFRKAKPEQIRAIS
                     QQATSPRHRGGERLAINLLSQSPVYSNQNGRFFEACPEDFSQFQNMDVAVSAFKLNQG
                     AIFVPHYNSKATFVVFVTDGYGYAQMACPHLSRQSQGSQSGRQDRREQEEESEEETFG
                     EFQQVKAPLSPGDVFVAPAGHAVTFFASKDQPLNAVAFGLNAQNNQRIFLAGKKNLVR
                     QMDSEAKELSFGVPSKLVDNIFNNPDESYFMSFSQQRQRRDERRGNPLASILDFARLF
                     "
ORIGIN

Alignment Scores:
Pred. No.:              3.68e-11        Length:         1867
Score:                  117.00          Matches:        23
Percent Similarity:     88.46%          Conservative:   0
Best Local Similarity:  88.46%          Mismatches:     3
Query Match:            95.90%          Indels:         0
DB:                     6               Gaps:           0

US-10-691-590-3 (1-26) x A20606 (1-1867)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db       1322 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 1381

Qy         21 AlaProLeu***ProGly 26
              |||||||||   ||||||
Db       1382 GCCCCATTGTCACCTGGT 1399

RESULT 2
TCCSV
LOCUS       TCCSV                   3318 bp    DNA     linear   PLN 03-DEC-1993
DEFINITION  T.cacao csv gene for seed vicilin.
ACCESSION   X62625 S38078
VERSION     X62625.1  GI:21910
KEYWORDS    csv gene; seed protein; vicilin.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 3318)
  AUTHORS   McHenry,L. and Fritz,P.J.
  TITLE     Comparison of the structure and nucleotide sequences of vicilin
            genes of cocoa and cotton raise questions about vicilin evolution
  JOURNAL   Plant Mol. Biol. 18 (6), 1173-1176 (1992)
  MEDLINE   92288309
   PUBMED   1600151
REFERENCE   2  (bases 1 to 3318)
  AUTHORS   McHenry,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1991) L. McHenry, Pennsylvania State University,
            111 Borland Lab, University Park, PA 16802, USA
COMMENT     See also x62626
            Overlap of sequenced fragments.
FEATURES             Location/Qualifiers
     source          1. .3318
                     /organism="Theobroma cacao"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3641"
                     /clone="pHD5P1.7, pHD5P1.2, pHD5E4.5"
                     /tissue_type="leaves"
                     /clone_lib="cocoa genomic library in lambda GEM II"
                     /dev_stage="mature tree"
     TATA_signal     543. .549
                     /note="csv gene"
     gene            610. .2581
                     /gene="csv"
     CDS             join(610. .1240,1331. .1506,1597. .1677,1780. .2064,
                     2177. .2581)
                     /gene="csv"
                     /codon_start=1
                     /product="vicilin"
                     /protein_id="CAA44493.1"
                     /db_xref="GI:21911"
                     /db_xref="GOA:Q43358"
                     /db_xref="UniProt/Swiss-Prot:Q43358"
                     /translation="MVISKSPFIVLIFSLLLSFALLCSGVSAYGRKQYERDPRQQYEQ
                     CQRRCESEATEEREQEQCEQRCEREYKEQQRQQEEELQRQYQQCQGRCQEQQQGQREQ
                     QQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQRNNPYYFPKRRSFQTRFRDEE
                     GNFKILQRFAENSPPLKGINDYRLAMFEANPNTFILPHHCDAEAIYFVTNGKGTITFV
                     THENKESYNVQRGTVVSVPAGSTVYVVSQDNQEKLTIAVLALPVNSPGKYELFFPAGN
                     NKPESYYGAFSYEVLETVFNTQREKLEEILEEQRGQKRQQGQQGMFRKAKPEQIRAIS
                     QQATSPRHRGGERLAINLLSQSPVYSNQNGRFFEACPEDFSQFQNMDVAVSAFKLNQG
                     AIFVPHYNSKATFVVFVTDGYGYAQMACPHLSRQSQGSQSGRQDRREQEEESEEETFG
                     EFQQVKAPLSPGDVFVAPAGHAVTFFASKDQPLNAVAFGLNAQNNQRIFLAGRPFFLN
                     HKQNTNVIKFTVKASAY"
     sig_peptide     610. .681
                     /gene="csv"
     mat_peptide     join(682. .1240,1331. .1506,1597. .1677,1780. .2064,
                     2177. .2578)
                     /gene="csv"
                     /product="vicilin"
     exon            <610. .1240
                     /gene="csv"
                     /number=1
     intron          1241. .1330
                     /gene="csv"
                     /number=1
     exon            1331. .1506
                     /gene="csv"
                     /number=2
     intron          1507. .1596
                     /gene="csv"
                     /number=2
     exon            1597. .1677
                     /gene="csv"
                     /number=3
     intron          1678. .1779
                     /gene="csv"
                     /number=3
     exon            1780. .2064
                     /gene="csv"
                     /number=4
     intron          2065. .2176
                     /gene="csv"
                     /number=4
     exon            2177. .>2582
                     /number=5
     polyA_signal    2856. .2861
ORIGIN

Alignment Scores:
Pred. No.:              6.82e-11        Length:         3318
Score:                  117.00          Matches:        23
Percent Similarity:     88.46%          Conservative:   0
Best Local Similarity:  88.46%          Mismatches:     3
Query Match:            95.90%          Indels:         0
DB:                     8               Gaps:           0

US-10-691-590-3 (1-26) x TCCSV (1-3318)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db       2312 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTCCAGCAGGTCAAA 2371

Qy         21 AlaProLeu***ProGly 26
              |||||||||   ||||||
Db       2372 GCCCCATTGTCACCTGGT 2389

RESULT 3
TCCSVSV
```

```
LOCUS       TCCSVSV                 1380 bp    mRNA    linear   PLN 06-DEC-1992
DEFINITION  T.cacao csv mRNA for seed vicilin.
ACCESSION   X62626
VERSION     X62626.1  GI:21912
KEYWORDS    csv gene; seed protein; vicilin.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 1380)
  AUTHORS   McHenry,L. and Fritz,P.J.
  TITLE     Comparison of the structure and nucleotide sequences of vicilin
            genes of cocoa and cotton raise questions about vicilin evolution
  JOURNAL   Plant Mol. Biol. 18 (6), 1173-1176 (1992)
  MEDLINE   92288309
   PUBMED   1600151
REFERENCE   2  (bases 1 to 1380)
  AUTHORS   McHenry,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1991) L. McHenry, Pennsylvania State University,
            111 Borland Lab, University Park, PA 16802, USA
COMMENT     See also x62625
            overlap of sequenced fragments, sequence partial.
FEATURES             Location/Qualifiers
     source          1. .1380
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /db_xref="taxon:3641"
                     /clone="pHT4C3"
                     /tissue_type="cotyledon"
                     /clone_lib="cocoa seed library in lambda gt10"
                     /dev_stage="130DAP"
     gene            1. .1380
                     /gene="csv"
     CDS             <1. .>1380
                     /gene="csv"
                     /codon_start=1
                     /product="vicilin"
                     /protein_id="CAA44494.1"
                     /db_xref="GI:21913"
                     /db_xref="GOA:Q43358"
                     /db_xref="UniProt/Swiss-Prot:Q43358"
                     /translation="PQIKSIANMVISKSPFIVLIFSLLLSFALLCSGVSAYGRKQYER
                     DPRQQYEQCQRRCESEATEEREQEQCEQRCEREYKEQQRQQEEELQRQYQQCQGRCQE
                     QQQGQREQQQCQRKCWEQYKEQERGEHENYHNHKKNRSEEEEGQQRNNPYYFPKRRSF
                     QTRFRDEEGNFKILQRFAENSPPLKGINDYRLAMFEANPNTFILPHHCDAEAIYFVTN
                     GKGTITFVTHENKESYNVQRGTVVSVPAGSTVYYVSQDNQEKLTIAVLALPVNSPGKY
                     ELFFPAGNNKPESYYGAFSYEVLETVFNTQREKLEEILEEQRGQKRQQGQQGMFRKAK
                     PEQIRAISQQATSPRHRGGERLAINLLSQSPVYSNQNGRFFEACPEDFSQFQNMDVAV
                     SAFKLNQGAIFVPHYNSKATFVVFVTDGYGYAQMACPHLSRQSQGSQSGRQDRREQEE
                     ESEEETFGEF"
     sig_peptide     <1. .24
                     /gene="csv"
     mat_peptide     25. .1380
                     /gene="csv"
                     /product="vicilin"
ORIGIN

Alignment Scores:
Pred. No.:                6.08e-05        Length:          1380
Score:                    82.00           Matches:         16
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              67.21%          Indels:          0
DB:                       8               Gaps:            0

US-10-691-590-3 (1-26) x TCCSVSV (1-1380)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe 16
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1333 AGAAGAGAACAAGAAGAAGAGTCAGAAGAGGAGACATTTGGAGAATTC 1380
```

```
RESULT 4
AC095192
LOCUS       AC095192               242889 bp    DNA     linear   HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-3D1, *** SEQUENCING IN PROGRESS ***,
            3 unordered pieces.
ACCESSION   AC095192
VERSION     AC095192.5  GI:24817922
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 242889)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 242889)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 242889)
```

```
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 9, 2002 this sequence version replaced gi:22773040.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GDCH
                Center clone name: CH230-3D1
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 219628 bases at least Q40
                Consensus quality: 223587 bases at least Q30
                Consensus quality: 226895 bases at least Q20
                Estimated insert size: 233065; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   154642: contig of 154642 bp in length
            *  154643   154742: gap of unknown length
            *  154743   239321: contig of 84579 bp in length
            *  239322   239421: gap of unknown length
            *  239422   242889: contig of 3468 bp in length.
FEATURES             Location/Qualifiers
     source          1. .242889
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-3D1"
ORIGIN

Alignment Scores:
Pred. No.:              12.6          Length:         242889
Score:                  66.00         Matches:        12
Percent Similarity:     65.22%        Conservative:   3
Best Local Similarity:  52.17%        Mismatches:     8
Query Match:            54.10%        Indels:         0
DB:                     2             Gaps:           0

US-10-691-590-3 (1-26) x AC095192 (1-242889)

Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal***AlaPro 22
              |||:::|||||||||   |||||||||      |||   |||      :::   :::|||
Db     233897 GAAGAAGAAGAAGAAGAAGAAGAAGTAGTCGGAACCTTTAAGAATATTGAGTCGCCC 233956

Qy         23 Leu***Pro 25
              |||   |||
```

```
Db     233957 CTCTTCCCC 233965

RESULT 5
AC136578/c
LOCUS       AC136578              252743 bp    DNA     linear   HTG 23-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-10P6, *** SEQUENCING IN PROGRESS ***,
            3 unordered pieces.
ACCESSION   AC136578
VERSION     AC136578.2  GI:25188799
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 252743)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnbunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 252743)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
```

```
REFERENCE   3  (bases 1 to 252743)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 23, 2002 this sequence version replaced gi:24635772.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GEAA
                Center clone name: CH230-10P6
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 225930 bases at least Q40
                Consensus quality: 229296 bases at least Q30
                Consensus quality: 231846 bases at least Q20
                Estimated insert size: 238263; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    17141: contig of 17141 bp in length
            *    17142    17241: gap of unknown length
            *    17242   242640: contig of 225399 bp in length
            *   242641   242740: gap of unknown length
            *   242741   252743: contig of 10003 bp in length.
FEATURES             Location/Qualifiers
     source          1. .252743
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-10P6"
     misc_feature    4537. .5487
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH309637"
     misc_feature    17242. .19226
                     /note="wgs_contig"
     misc_feature    complement(220741. .221405)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH309638"
     misc_feature    241487. .242640
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    242741. .244272
                     /note="wgs_end_extension
                     clone_end:Sp6"
ORIGIN

Alignment Scores:
Pred. No.:              13.1          Length:         252743
Score:                  66.00         Matches:        12
Percent Similarity:     65.22%        Conservative:   3
Best Local Similarity:  52.17%        Mismatches:     8
Query Match:            54.10%        Indels:         0
DB:                     2             Gaps:           0

US-10-691-590-3 (1-26) x AC136578 (1-252743)

Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal***AlaPro 22
              |||:::|||||||||   |||||||||      |||   |||      :::   :::|||
Db      49987 GAAGAAGAAGAAGAAGAAGAAGAAGTAGTCGGAACCTTTAAGAATATTGAGTCGCCC 49928

Qy         23 Leu***Pro 25
              |||   |||
Db      49927 CTCTTCCCC 49919

RESULT 6
AC010649/c
LOCUS       AC010649               37492 bp    DNA     linear   PRI 30-AUG-2001
DEFINITION  Homo sapiens chromosome 19 clone LLNLR-304C3, complete sequence.
ACCESSION   AC010649
VERSION     AC010649.8  GI:15375139
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 37492)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 37492)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 37492)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 37492)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Aug 30, 2001 this sequence version replaced gi:14916144.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.5% of Sequence;
            Estimated Total Number of Errors is 0.1.
FEATURES             Location/Qualifiers
     source          1. .37492
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="LLNLR-304C3"
ORIGIN

Alignment Scores:
Pred. No.:              2.09          Length:         37492
Score:                  65.50         Matches:        14
Percent Similarity:     59.26%        Conservative:   2
Best Local Similarity:  51.85%        Mismatches:     8
Query Match:            53.69%        Indels:         3
```

```
DB:                    9              Gaps:          1

US-10-691-590-3 (1-26) x AC010649 (1-37492)

Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***Gln---------Val 19
              |||:::|||||||||   |||||||||   ||||||   |||   |||          :::
Db      19577 GAGGAGGAGGAAGAAGAGGAGGAAGAATTATTTGGCAATTTCTCACAAAGGTCAAACATG 19518

Qy         20 ***AlaProLeu***ProGly 26
                  |||      ||||||
Db      19517 GAGCCACCATTTGACCCAGGA 19497

RESULT 7
AC102322
LOCUS       AC102322               171432 bp    DNA     linear   HTG 21-AUG-2002
DEFINITION  Mus musculus clone RP24-462C5, WORKING DRAFT SEQUENCE, 29 unordered
            pieces.
ACCESSION   AC102322
VERSION     AC102322.2  GI:22381079
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 171432)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-462C5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171432)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 171432)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 21, 2002 this sequence version replaced gi:17061408.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L18346
                Center clone name: 462_C_5
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 158744 bases at least Q40
                Consensus quality: 162950 bases at least Q30
                Consensus quality: 164899 bases at least Q20
                Insert size: 176000; agarose-fp
                Insert size: 168632; sum-of-contigs
                Quality coverage: 5.3 in Q20 bases; agarose-fp
                Quality coverage: 5.6 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 29 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      681: contig of 681 bp in length
            *      682      781: gap of 100 bp
            *      782     1476: contig of 695 bp in length
            *     1477     1576: gap of 100 bp
            *     1577     2062: contig of 486 bp in length
            *     2063     2162: gap of 100 bp
            *     2163     2595: contig of 433 bp in length
            *     2596     2695: gap of 100 bp
            *     2696     3398: contig of 703 bp in length
            *     3399     3498: gap of 100 bp
            *     3499     4884: contig of 1386 bp in length
            *     4885     4984: gap of 100 bp
            *     4985     5587: contig of 603 bp in length
            *     5588     5687: gap of 100 bp
            *     5688     6845: contig of 1158 bp in length
            *     6846     6945: gap of 100 bp
            *     6946     7876: contig of 931 bp in length
            *     7877     7976: gap of 100 bp
            *     7977     9201: contig of 1225 bp in length
            *     9202     9301: gap of 100 bp
            *     9302    11037: contig of 1736 bp in length
            *    11038    11137: gap of 100 bp
            *    11138    13003: contig of 1866 bp in length
            *    13004    13103: gap of 100 bp
            *    13104    15285: contig of 2182 bp in length
            *    15286    15385: gap of 100 bp
            *    15386    17319: contig of 1934 bp in length
            *    17320    17419: gap of 100 bp
            *    17420    19929: contig of 2510 bp in length
            *    19930    20029: gap of 100 bp
            *    20030    22691: contig of 2662 bp in length
            *    22692    22791: gap of 100 bp
            *    22792    25794: contig of 3003 bp in length
            *    25795    25894: gap of 100 bp
            *    25895    28135: contig of 2241 bp in length
            *    28136    28235: gap of 100 bp
            *    28236    31138: contig of 2903 bp in length
            *    31139    31238: gap of 100 bp
```

```
            *    31239    34733: contig of 3495 bp in length
            *    34734    34833: gap of 100 bp
            *    34834    40484: contig of 5651 bp in length
            *    40485    40584: gap of 100 bp
            *    40585    45922: contig of 5338 bp in length
            *    45923    46022: gap of 100 bp
            *    46023    51562: contig of 5540 bp in length
            *    51563    51662: gap of 100 bp
            *    51663    58135: contig of 6473 bp in length
            *    58136    58235: gap of 100 bp
            *    58236    68445: contig of 10210 bp in length
            *    68446    68545: gap of 100 bp
            *    68546    79090: contig of 10545 bp in length
            *    79091    79190: gap of 100 bp
            *    79191    92907: contig of 13717 bp in length
            *    92908    93007: gap of 100 bp
            *    93008   110376: contig of 17369 bp in length
            *   110377   110476: gap of 100 bp
            *   110477   171432: contig of 60956 bp in length.
FEATURES             Location/Qualifiers
     source          1. .171432
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-462C5"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .681
                     /note="assembly_fragment"
     misc_feature    782. .1476
                     /note="assembly_fragment"
     misc_feature    1577. .2062
                     /note="assembly_fragment"
     misc_feature    2163. .2595
                     /note="assembly_fragment"
     misc_feature    2696. .3398
                     /note="assembly_fragment"
     misc_feature    3499. .4884
                     /note="assembly_fragment"
     misc_feature    4985. .5587
                     /note="assembly_fragment"
     misc_feature    5688. .6845
                     /note="assembly_fragment"
     misc_feature    6946. .7876
                     /note="assembly_fragment"
     misc_feature    7977. .9201
                     /note="assembly_fragment"
     misc_feature    9302. .11037
                     /note="assembly_fragment"
     misc_feature    11138. .13003
                     /note="assembly_fragment"
     misc_feature    13104. .15285
                     /note="assembly_fragment"
     misc_feature    15386. .17319
                     /note="assembly_fragment"
     misc_feature    17420. .19929
                     /note="assembly_fragment"
     misc_feature    20030. .22691
                     /note="assembly_fragment"
     misc_feature    22792. .25794
                     /note="assembly_fragment"
     misc_feature    25895. .28135
                     /note="assembly_fragment"
     misc_feature    28236. .31138
                     /note="assembly_fragment"
     misc_feature    31239. .34733
                     /note="assembly_fragment"
     misc_feature    34834. .40484
                     /note="assembly_fragment"

Alignment Scores:
Pred. No.:              70.1          Length:          171432
Score:                  61.00         Matches:         12
Percent Similarity:     65.22%        Conservative:    3
Best Local Similarity:  52.17%        Mismatches:      8
Query Match:            50.00%        Indels:          0
DB:                     2             Gaps:            0

US-10-691-590-3 (1-26) x AC102322 (1-171432)

Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal***AlaPro 22
              |||:::|||||||||  |||||||||  |||||||||      :::        |||
Db     112717 GAAGAAGAAGAAGAAGAAGAAGAAGAATTTGGAGAGCTGGCACGGTAGTGCATGCCT 112776

Qy         23 Leu***Pro 25
              :::  |||
Db     112777 GTGATCCCA 112785

RESULT 8
AC121397
LOCUS       AC121397               214464 bp    DNA     linear   HTG 15-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-83K4, *** SEQUENCING IN PROGRESS ***,
            2 unordered pieces.
ACCESSION   AC121397
VERSION     AC121397.4  GI:25008414
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 214464)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
```

```
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 214464)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 214464)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 15, 2002 this sequence version replaced gi:23267576.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GWVD
                Center clone name: CH230-83K4
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 194421 bases at least Q40
                Consensus quality: 197717 bases at least Q30
                Consensus quality: 199996 bases at least Q20
                Estimated insert size: 201489; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   212804: contig of 212804 bp in length
            *  212805   212904: gap of unknown length
            *  212905   214464: contig of 1560 bp in length.
FEATURES            Location/Qualifiers
     source         1. .214464
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-83K4"
     misc_feature   1. .1268
                    /note="wgs_end_extension
                    clone_end:T7"
     misc_feature   364. .1268
                    /note="clone_boundary
                    clone_end:T7
                    site:
                    end_sequence:BH274376"
     misc_feature   complement(211702. .212641)
                    /note="clone_boundary
                    clone_end:Sp6
                    site:
                    end_sequence:BH274378"
ORIGIN

Alignment Scores:
Pred. No.:                136             Length:         214464
Score:                    60.00           Matches:        14
Percent Similarity:       57.69%          Conservative:   1
Best Local Similarity:    53.85%          Mismatches:     11
Query Match:              49.18%          Indels:         0
DB:                       2               Gaps:           0

US-10-691-590-3 (1-26) x AC121397 (1-214464)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              |||||||||:::|||||||||   |||||||||          |||          |||
Db     197123 AGGAGAGAGGAGGAAGAGGAGGAGGAAGAGGAGGACGCTACTGAGACATTATCTGTGACT 197182

Qy         21 AlaProLeu***ProGly 26
                  |||   ||||||
Db     197183 AAGATCCTGGGGCCAGGA 197200

RESULT 9
AC097079
LOCUS       AC097079              236863 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-85G9, WORKING DRAFT SEQUENCE.
ACCESSION   AC097079
VERSION     AC097079.7  GI:30522186
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 236863)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
```

```
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 236863)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 236863)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:23264495.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GGXE
                Center clone name: CH230-85G9
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 226010 bases at least Q40
                Consensus quality: 228329 bases at least Q30
                Consensus quality: 229882 bases at least Q20
                Estimated insert size: 237449; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   236863: contig of 236863 bp in length.
FEATURES            Location/Qualifiers
     source         1. .236863
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-85G9"
     misc_feature   340. .1181
                    /note="clone_boundary
                    clone_end:T7
                    site:EcoRI
                    end_sequence:BH305657"
     misc_feature   173038. .173892
                    /note="clone_boundary
                    clone_end:Sp6
                    site:EcoRI
                    end_sequence:BH305659"
     misc_feature   230193. .231771
                    /note="wgs_end_extension
                    clone_end:Sp6"
     misc_feature   233492. .236863
                    /note="wgs_end_extension
                    clone_end:Sp6"
ORIGIN

Alignment Scores:
Pred. No.:              151             Length:         236863
Score:                  60.00           Matches:        14
Percent Similarity:     57.69%          Conservative:   1
Best Local Similarity:  53.85%          Mismatches:     11
Query Match:            49.18%          Indels:         0
DB:                     2               Gaps:           0

US-10-691-590-3 (1-26) x AC097079 (1-236863)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              ||||||||::::||||||||   ||||||||         |||         |||
Db      25787 AGGAGAGAGGAGGAAGAGGAGGAGGAAGAGGAGGACGCTACTGAGACATTATCTGTGACT 25846

Qy         21 AlaProLeu***ProGly 26
                  |||   ||||||
Db      25847 AAGATCCTGGGGCCAGGA 25864

RESULT 10
AC115994
LOCUS       AC115994               242913 bp    DNA     linear   PRI 01-MAR-2003
DEFINITION  Homo sapiens chromosome 17, clone CTD-2535L24, complete sequence.
ACCESSION   AC115994
VERSION     AC115994.24  GI:28626868
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 242913)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone CTD-2535L24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 242913)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
```

```
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek;L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 242913)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 242913)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 1, 2003 this sequence version replaced gi:28173200.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L26130
                Center clone name: 2535_L_24
            --------------.
FEATURES            Location/Qualifiers
     source         1. .242913
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="17"
                    /map="17"
                    /clone="CTD-2535L24"
                    /clone_lib="CITD2 Human BAC"
     repeat_region  complement(1. .80)
                    /rpt_family="AluJ/FLAM"
     repeat_region  complement(250. .553)
                    /rpt_family="AluY"
     repeat_region  complement(1209. .1510)
                    /rpt_family="AluSq"
     repeat_region  1595. .1903
                    /rpt_family="AluSx"
     repeat_region  complement(2033. .2195)
                    /rpt_family="MLT1F2"
     repeat_region  complement(2233. .2425)
                    /rpt_family="MLT1F2"
     repeat_region  complement(2601. .2650)
                    /rpt_family="L2"
     repeat_region  2859. .3173
                    /rpt_family="AluSq"
     repeat_region  complement(3546. .3842)
                    /rpt_family="AluSg"
     repeat_region  4125. .4432
                    /rpt_family="AluSg"
     repeat_region  4448. .4471
                    /rpt_family="AT_rich"
     repeat_region  complement(4765. .5063)
                    /rpt_family="AluSx"
     repeat_region  5482. .5595
                    /rpt_family="MIR3"
     repeat_region  complement(5596. .5905)
                    /rpt_family="AluSq"
     repeat_region  5906. .5961
                    /rpt_family="MIR3"
     repeat_region  complement(6188. .6311)
                    /rpt_family="MIR"
     repeat_region  complement(7068. .7373)
                    /rpt_family="AluSp"
     repeat_region  7692. .7737
                    /rpt_family="(TG)n"
     repeat_region  complement(8354. .8637)
                    /rpt_family="AluSx"
     repeat_region  complement(8853. .9158)
                    /rpt_family="AluSx"
     repeat_region  10104. .10140
                    /rpt_family="(CGGGG)n"
     repeat_region  10153. .10182
                    /rpt_family="GC_rich"
     repeat_region  10328. .10358
                    /rpt_family="GC_rich"
     repeat_region  10435. .10581
                    /rpt_family="(CGGGG)n"
     repeat_region  10723. .10750
                    /rpt_family="GC_rich"
     repeat_region  10882. .10950
                    /rpt_family="GC_rich"
     repeat_region  11388. .11463
                    /rpt_family="C-rich"
     repeat_region  11566. .11603
                    /rpt_family="GC_rich"
     repeat_region  11691. .11732
                    /rpt_family="GC_rich"
     repeat_region  16081. .16232
                    /rpt_family="MIR"
     repeat_region  17557. .17600
                    /rpt_family="(GGA)n"
     repeat_region  complement(18583. .18760)
                    /rpt_family="L2"
     repeat_region  complement(18771. .19079)
                    /rpt_family="AluSp"
     repeat_region  complement(19354. .19677)
                    /rpt_family="L3b"
     repeat_region  complement(19763. .19879)
```

```
                     /rpt_family="MIR"
     repeat_region   20801. .20889
                     /rpt_family="MIR"
     repeat_region   complement(23168. .23264)
                     /rpt_family="MER5B"
     repeat_region   23265. .23383
                     /rpt_family="MARNA"
     repeat_region   complement(26163. .26418)
                     /rpt_family="AluSc"
     repeat_region   28178. .28292
                     /rpt_family="MIR"
     repeat_region   29267. .29563
                     /rpt_family="L2"
     unsure          30337. .30342
                     /note="<30 qual SNGL region"
     repeat_region   complement(30583. .30648)
                     /rpt_family="MSTB1"
     repeat_region   30649. .30948
                     /rpt_family="AluSx"
     repeat_region   complement(30949. .31273)
                     /rpt_family="MSTB1"
     repeat_region   31354. .31554

Alignment Scores:
Pred. No.:              155          Length:        242913
Score:                  60.00        Matches:       11
Percent Similarity:     76.19%       Conservative:  5
Best Local Similarity:  52.38%       Mismatches:    5
Query Match:            49.18%       Indels:        0
DB:                     9            Gaps:          0

US-10-691-590-3 (1-26) x AC115994 (1-242913)

Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal***AlaPro 22
              |||:::|||||||||   ||||||:::   :::|||:::|||   ||||||   :::
Db      17575 GAGGAAGAGGAGGAGAAGGAGGAGGACCCCTATGGTCAGTTCTGTCAGGTCAGTTCTGTG 17634

Qy         23 Leu 23
              |||
Db      17635 CTC 17637

RESULT 11
BC069795
LOCUS       BC069795                2332 bp    mRNA    linear   PRI 30-JUN-2004
DEFINITION  Homo sapiens histidine rich calcium binding protein, mRNA (cDNA
            clone MGC:97245 IMAGE:7262494), complete cds.
ACCESSION   BC069795
VERSION     BC069795.1  GI:46854901
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2332)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2332)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-2004) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Baylor Human Genome Sequencing Center
            cDNA Library Preparation: Baylor Human Genome Sequencing Center
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRBR Plate: 5 Row: b Column: 8.
FEATURES              Location/Qualifiers
     source           1. .2332
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="MGC:97245 IMAGE:7262494"
                      /tissue_type="PCR rescued clones"
                      /clone_lib="NIH_MGC_244"
                      /note="Vector: pPCR-Script Amp SK(+)"
     gene             1. .2332
                      /gene="HRC"
                      /db_xref="LocusID:3270"
                      /db_xref="MIM:142705"
     CDS              171. .2270
                      /gene="HRC"
                      /codon_start=1
                      /product="histidine-rich calcium-binding protein,
                      precursor"
                      /protein_id="AAH69795.1"
                      /db_xref="GI:46854902"
                      /db_xref="LocusID:3270"
                      /db_xref="MIM:142705"
                      /translation="MGHHRPWLHASVLWAGVASLLLPPAMTQQLRGDGLGFRNRNNST
                      GVAGLSEEASAELRHHLHSPRDHPDENKDVSTENGHHFWSHPDREKEDEDVSKEYGHL
                      LPGHRSQDHKVGDEGVSGEEVFAEHGGQARGHRGHGSEDTEDSAEHRHHLPSHRSHSH
                      QDEDEDEVVSSEHHHHILRHGHRGHDGEDDEGEEEEEEEEEEEEEASTEYGHQAHRHRG
                      HGSEEDEDVSDGHHHHGPSHRHQGHEEDDDDDDDDDDDDDDDDDVSIEYRHQAHRHQGH
                      GIEEDEDVSDGHHHRDPSHRHRSHEEDDNDDDDVSTEYGHQAHRHQDHRKEEVEAVSG
                      EHHHHVPDHRHQGHRDEEEDEDVSTERWHQGPQHVHHGLVDEEEEEEEITVQFGHYVA
                      SHQPRGHKSDEEDFQDEYKTEVPHHHHHRVPREEDEEVSAELGHQAPSHRQSHQDEET
                      GHGQRGSIKEMSHHPPGHTVVKDRSHLRKDDSEEEKEKEEDPGSHEEDDESSEQGEKG
                      THHGSRDQEDEEDEEEGHGLSLNQEEEEEEDKEEEEEEEDEERREERAEVGAPLSPDH
                      SEEEEEEEEGLEEDEPRFTIIPNPLDRREEAGGASSEEESGEDTGPQDAQEYGNYQPG
                      SLCGYCSFCNRCTECESCHCDEENMGEHCDQCQHCQFCYLCPLVCETVCAPGSYVDYF
                      SSSLYQALADMLETPEP"
ORIGIN

Alignment Scores:
Pred. No.:              1.61         Length:        2332
Score:                  59.00        Matches:       12
Percent Similarity:     64.00%       Conservative:  4
Best Local Similarity:  48.00%       Mismatches:    9
Query Match:            48.36%       Indels:        0
DB:                     9            Gaps:          0
```

```
US-10-691-590-3 (1-26) x BC069795 (1-2332)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              :::   |||:::|||||||||   :::||||||        |||      :::|||
Db       1788 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 1847

Qy         21 AlaProLeu***Pro 25
              |||||||||   |||
Db       1848 GCCCCACTGAGCCCA 1862

RESULT 12
BC069802
LOCUS       BC069802               2332 bp    mRNA    linear   PRI 30-JUN-2004
DEFINITION  Homo sapiens histidine rich calcium binding protein, mRNA (cDNA
            clone MGC:97257 IMAGE:7262506), complete cds.
ACCESSION   BC069802
VERSION     BC069802.1  GI:46854908
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2332)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2332)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-2004) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Baylor Human Genome Sequencing Center
            cDNA Library Preparation: Baylor Human Genome Sequencing Center
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRBR Plate: 5 Row: c Column: 8.
FEATURES             Location/Qualifiers
     source          1. .2332
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:97257 IMAGE:7262506"
                     /tissue_type="PCR rescued clones"
                     /clone_lib="NIH_MGC_244"
                     /note="Vector: pPCR-Script Amp SK(+)"
     gene            1. .2332
                     /gene="HRC"
                     /db_xref="LocusID:3270"
                     /db_xref="MIM:142705"
     CDS             171. .2270
                     /gene="HRC"
                     /codon_start=1
                     /product="histidine-rich calcium-binding protein,
                     precursor"
                     /protein_id="AAH69802.1"
                     /db_xref="GI:46854909"
                     /db_xref="LocusID:3270"
                     /db_xref="MIM:142705"
                     /translation="MGHHRPWLHASVLWAGVASLLLPPAMTQQLRGDGLGFRNRNNST
                     GVAGLSEEASAELRHHLHSPRDHPDENKDVSTENGHHFWSHPDREKEDEDVSKEYGHL
                     LPGHRSQDHKVGDEGVSGEEVFAEHGGQARGHRGHGSEDTEDSAEHRHHLPSHRSHSH
                     QDEDEDEVVSSEHHHHILRHGHRGHDGEDDEGEEEEEEEEEEEEASTEYGHQAHRHRG
                     HGSEEDEDVSDGHHHHGPSHRHQGHEEDDDDDDDDDDDDDDDDVSIEYRHQAHRHQGH
                     GIEEDEDVSDGHHHRDPSHRHRSHEEDDNDDDDVSTEYGHQAHRHQDHRKEEVEAVSG
                     EHHHHVPDHRHQGHRDEEEDEDVSTERWHQGPQHVHHGLVDEEEEEEEITVQFGHYVA
                     SHQPRGHKSDEEDFQDEYKTEVPHHHHHRVPREEDEEVSAELGHQAPSHRQSHQDEET
                     GHGQRGSIKEMSHHPPGHTVVKDRSHLRKDDSEEEKEKEEDPGSHEEDDESSEQGEKG
                     THHGSRDQEDEEDEEEGHGLSLNQEEEEEEDKEEEEEEEDEERREERAEVGAPLSPDH
                     SEEEEEEEEGLEEDEPRFTIIPNPLDRREEAGGASSEEESGEDTGPQDAQEYGNYQPG
                     SLCGYCSFCNRCTECESCHCDEENMGEHCDQCQHCQFCYLCPLVCETVCAPGSYVDYF
                     SSSLYQALADMLETPEP"
ORIGIN

Alignment Scores:
Pred. No.:              1.61          Length:         2332
Score:                  59.00         Matches:        12
Percent Similarity:     64.00%        Conservative:   4
Best Local Similarity:  48.00%        Mismatches:     9
Query Match:            48.36%        Indels:         0
DB:                     9             Gaps:           0

US-10-691-590-3 (1-26) x BC069802 (1-2332)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              :::   |||:::|||||||||   :::||||||        |||      :::|||
Db       1788 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 1847

Qy         21 AlaProLeu***Pro 25
              |||||||||   |||
Db       1848 GCCCCACTGAGCCCA 1862

RESULT 13
CQ715312
LOCUS       CQ715312               2365 bp    DNA     linear   PAT 03-FEB-2004
DEFINITION  Sequence 1246 from Patent WO02068579.
ACCESSION   CQ715312
VERSION     CQ715312.1  GI:42276169
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
  TITLE     Kits, such as nucleic acid arrays, comprising a majority of
            humanexons or transcripts, for detecting expression and other uses
            thereof
  JOURNAL   Patent: WO 02068579-A 1246 06-SEP-2002;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .2365
                     /organism="Homo sapiens"
```

```
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:                1.63          Length:          2365
Score:                    59.00         Matches:         12
Percent Similarity:       64.00%        Conservative:    4
Best Local Similarity:    48.00%        Mismatches:      9
Query Match:              48.36%        Indels:          0
DB:                       6             Gaps:            0

US-10-691-590-3 (1-26) x CQ715312 (1-2365)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              :::   |||:::|||||||||   :::||||||         |||      :::|||
Db       1788 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 1847

Qy         21 AlaProLeu***Pro 25
              |||||||||   |||
Db       1848 GCCCCACTGAGCCCA 1862

RESULT 14
HUMHCPB
LOCUS       HUMHCPB                 2365 bp    mRNA    linear   PRI 14-FEB-1996
DEFINITION  Human histidine-rich calcium binding protein (HRC) mRNA, complete
            cds.
ACCESSION   M60052 J04784 J04785 J04786
VERSION     M60052.1  GI:183918
KEYWORDS    histidine-rich calcium binding protein; luminal sarcoplasmic
            reticulum protein.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2365)
  AUTHORS   Hofmann,S.L., Topham,M., Hsieh,C.L. and Francke,U.
  TITLE     cDNA and genomic cloning of HRC, a human sarcoplasmic reticulum
            protein, and localization of the gene to human chromosome 19 and
            mouse chromosome 7
  JOURNAL   Genomics 9 (4), 656-669 (1991)
  MEDLINE   91244309
   PUBMED   2037293
COMMENT     Original source text: Homo sapiens adult skeletal muscle cDNA to
            mRNA.
FEATURES             Location/Qualifiers
     source          1. .2365
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /map="19q13.2-q13.3"
                     /tissue_type="skeletal muscle"
                     /dev_stage="adult"
     gene            1. .2365
                     /gene="HRC"
     CDS             171. .2270
                     /gene="HRC"
                     /codon_start=1
                     /product="histidine-rich calcium binding protein"
                     /protein_id="AAA88071.1"
                     /db_xref="GI:183919"
                     /db_xref="GDB:G00-126-369"
                     /translation="MGHHRPWLHASVLWAGVASLLLPPAMTQQLRGDGLGFRNRNNST
                     GVAGLSEEASAELRHHLHSPRDHPDENKDVSTENGHHFWSHPDREKEDEDVSKEYGHL
                     LPGHRSQDHKVGDEGVSGEEVFAEHGGQARGHRGHGSEDTEDSAEHRHHLPSHRSHSH
                     QDEDEDEVVSSEHHHHILRHGHRGHDGEDDEGEEEEEEEEEEEASTEYGHQAHRHRG
                     HGSEEDEDVSDGHHHHGPSHRHQGHEEDDDDDDDDDDDDDDDVSIEYRHQAHRHQGH
                     GIEEDEDVSDGHHHRDPSHRHRSHEEDDNDDDVSTEYGHQAHRHQDHRKEEVEAVSG
                     EHHHHVPDHRHQGHRDEEEDEDVSTERWHQGPQHVHHGLVDEEEEEEEITVQFGHYVA
                     SHQPRGHKSDEEDFQDEYKTEVPHHHHHRVPREEDEEVSAELGHQAPSHRQSHQDEET
                     GHGQRGSIKEMSHHPPGHTVVKDRSHLRKDDSEEEKEKEEDPGSHEEDDESSEQGEKG
                     THHGSRDQEDEEDEEEGHGLSLNQEEEEEEDKEEEEEEEDEERREERAEVGAPLSPDH
                     SEEEEEEEEGLEEDEPRFTIIPNPLDRREEAGGASSEEESGEDTGPQDAQEYGNYQPG
                     SLCGYCSFCNRCTECESCHCDEENMGEHCDQCQHCQFCYLCPLVCETVCAPGSYVDYF
                     SSSLYQALADMLETPEP"
     sig_peptide     171. .254
                     /gene="HRC"
                     /note="G00-126-369"
     mat_peptide     255. .2267
                     /gene="HRC"
                     /product="histidine-rich calcium binding protein"
                     /note="G00-126-369"
ORIGIN      human chromosome 19q13.3 and mouse chromosome 7.

Alignment Scores:
Pred. No.:                1.63          Length:          2365
Score:                    59.00         Matches:         12
Percent Similarity:       64.00%        Conservative:    4
Best Local Similarity:    48.00%        Mismatches:      9
Query Match:              48.36%        Indels:          0
DB:                       9             Gaps:            0

US-10-691-590-3 (1-26) x HUMHCPB (1-2365)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              :::   |||:::|||||||||   :::||||||         |||      :::|||
Db       1788 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 1847

Qy         21 AlaProLeu***Pro 25
              |||||||||   |||
Db       1848 GCCCCACTGAGCCCA 1862

RESULT 15
AC145716/c
LOCUS       AC145716                42003 bp    DNA     linear   HTG 22-JUL-2003
DEFINITION  Homo sapiens chromosome 19 clone XXfos-85506H7, WORKING DRAFT
            SEQUENCE, 4 unordered pieces.
ACCESSION   AC145716
VERSION     AC145716.1  GI:33112707
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 42003)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 42003)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUL-2003) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 3436408
            Center clone name: WF-85506H7
            -------
            Summary Statistics
            Consensus quality: 40613 bases at least Q40
            Consensus quality: 40943 bases at least Q30
            Consensus quality: 41232 bases at least Q20
            Estimated insert size: 40000; agarose-fp estimation
            Estimated insert size: 41703; sum-of-contigs estimation
            Quality coverage:  8.78 in Q20 bases; agarose-fp estimation
            Quality coverage:  8.42 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
```

```
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1672: contig of 1672 bp in length
            *     1673     1772: gap of unknown length
            *     1773     9634: contig of 7862 bp in length
            *     9635     9734: gap of unknown length
            *     9735    17785: contig of 8051 bp in length
            *    17786    17885: gap of unknown length
            *    17886    42003: contig of 24118 bp in length.
FEATURES             Location/Qualifiers
     source          1. .42003
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="XXfos-85506H7"
ORIGIN

Alignment Scores:
Pred. No.:              35.8          Length:         42003
Score:                  59.00         Matches:        12
Percent Similarity:     64.00%        Conservative:   4
Best Local Similarity:  48.00%        Mismatches:     9
Query Match:            48.36%        Indels:         0
DB:                     2             Gaps:           0

US-10-691-590-3 (1-26) x AC145716 (1-42003)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal*** 20
              :::   |||:::|||||||||   :::||||||        |||     :::|||
Db      20968 AAGGAGGAGGAGGAGGAGGAAGAAGACGAGGAGAGGAGGGAAGAGAGGGCTGAGGTTGGG 20909

Qy         21 AlaProLeu***Pro 25
              |||||||||   |||
Db      20908 GCCCCACTGAGCCCA 20894


Search completed: November  5, 2005, 04:43:00
Job time : 1390.78 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 5, 2005, 02:54:27 ; Search time 900.835 Seconds
(without alignments)
817.035 Million cell updates/sec

Title: US-10-691-590-1
Perfect score: 458
Sequence: 1 RREQEEESEEETFGEFQQVK.....FLNHKQNTNVIKFTVKASAY 89

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10691590/runat_02112005_120023_13233/app_query.fasta_1.462
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10691590_@CGN_1_1_582_@runat_02112005_120023_13233
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
23: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
24: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq2:*
25: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
26: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
27: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
28: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	159	34.7	562	20	US-10-021-323-762	Sequence 762, App
	2	156.5	34.2	2032	10	US-09-731-375A-1	Sequence 1, Appli
	3	156.5	34.2	2032	15	US-10-228-806-1	Sequence 1, Appli
	4	156.5	34.2	2032	18	US-10-100-303A-5	Sequence 5, Appli
	5	156.5	34.2	2032	21	US-10-728-323-1	Sequence 1, Appli
	6	156.5	34.2	2032	21	US-10-728-051-1	Sequence 1, Appli
	7	156.5	34.2	2032	22	US-10-899-551-1	Sequence 1, Appli
	8	156.5	34.2	2032	22	US-10-958-324-5	Sequence 5, Appli
c	9	154.5	33.7	498	20	US-10-021-323-2828	Sequence 2828, Ap
	10	151.5	33.1	2003	20	US-10-437-963-43667	Sequence 43667, A
	11	151	33.0	1949	18	US-10-100-303A-6	Sequence 6, Appli
	12	145.5	31.8	1587	19	US-10-425-114-7095	Sequence 7095, Ap
	13	145.5	31.8	1933	19	US-10-425-114-23659	Sequence 23659, A
	14	145.5	31.8	2001	19	US-10-425-114-3415	Sequence 3415, Ap
	15	145.5	31.8	2926	21	US-10-425-115-117618	Sequence 117618,
	16	143	31.2	614	19	US-10-425-114-8131	Sequence 8131, Ap
	17	143	31.2	720	19	US-10-425-114-15167	Sequence 15167, A
	18	143	31.2	1037	19	US-10-425-114-14858	Sequence 14858, A
	19	143	31.2	1107	19	US-10-425-114-8117	Sequence 8117, Ap
	20	143	31.2	1656	19	US-10-425-114-9571	Sequence 9571, Ap
c	21	143	31.2	5642	19	US-10-424-599-10364	Sequence 10364, A
	22	142	31.0	1920	9	US-09-758-652-2	Sequence 2, Appli
	23	142	31.0	1920	19	US-10-684-651-2	Sequence 2, Appli
	24	142	31.0	1920	20	US-10-757-074-2	Sequence 2, Appli
	25	142	31.0	1920	20	US-10-757-155-2	Sequence 2, Appli
	26	142	31.0	1920	20	US-10-757-667-2	Sequence 2, Appli
	27	140.5	30.7	2171	9	US-09-331-631A-4	Sequence 4, Appli
	28	140.5	30.7	2171	17	US-10-147-095-4	Sequence 4, Appli
	29	140	30.6	1254	18	US-10-245-227B-15	Sequence 15, Appl
	30	140	30.6	1278	18	US-10-245-227B-13	Sequence 13, Appl
	31	140	30.6	1320	9	US-09-758-652-3	Sequence 3, Appli
	32	140	30.6	1320	18	US-10-245-227B-3	Sequence 3, Appli
	33	140	30.6	1320	19	US-10-684-651-3	Sequence 3, Appli
	34	140	30.6	1320	20	US-10-757-074-3	Sequence 3, Appli
	35	140	30.6	1320	20	US-10-757-155-3	Sequence 3, Appli
	36	140	30.6	1320	20	US-10-757-667-3	Sequence 3, Appli
	37	137	29.9	342	19	US-10-425-114-8154	Sequence 8154, Ap
	38	137	29.9	536	19	US-10-425-114-7591	Sequence 7591, Ap
	39	137	29.9	641	19	US-10-425-114-8571	Sequence 8571, Ap
	40	137	29.9	673	19	US-10-425-114-7653	Sequence 7653, Ap
	41	137	29.9	863	19	US-10-425-114-7590	Sequence 7590, Ap
	42	137	29.9	880	19	US-10-425-114-7667	Sequence 7667, Ap
	43	137	29.9	934	19	US-10-425-114-6929	Sequence 6929, Ap
	44	137	29.9	951	19	US-10-425-114-8579	Sequence 8579, Ap
	45	137	29.9	1011	19	US-10-425-114-8555	Sequence 8555, Ap

ALIGNMENTS

RESULT 1
US-10-021-323-762/c
; Sequence 762, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021,323
; CURRENT FILING DATE: 2001-12-12

```
;  PRIOR APPLICATION NUMBER: US 60/255, 619
;  PRIOR FILING DATE: 2000-12-14
;  NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 762
;   LENGTH: 562
;   TYPE: DNA
;   ORGANISM: Gossypium hirsutum
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3825-012-Q1-N6-A5
US-10-021-323-762

Alignment Scores:
Pred. No.:              5.29e-14      Length:        562
Score:                  159.00        Matches:       33
Percent Similarity:     64.47%        Conservative:  16
Best Local Similarity:  43.42%        Mismatches:    19
Query Match:            34.72%        Indels:        8
DB:                     20            Gaps:          2

US-10-691-590-1 (1-89) x US-10-021-323-762 (1-562)

Qy          3 GluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLysAlaPro 22
              |||   :::::::::   ||||||          |||:::::::::::::::::::::
Db        556 GAAGACCAAGACCAAGAAGAGGAAAGGAGAAGTGGCCAATACAGGAAAATAAGGTCCCGG 497

Qy         23 LeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPheAlaSer 42
              ||||||   ||||||:::||||||   ||||||         |||||||||   ||||||
Db        496 TTGTCGCGTGGAGACATATTTGTAGTCCCAGCAAACTTTCCAGTTACATTCGTTGCATCC 437

Qy         43 LysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsn------------ 58
              ::::::|||   |||             |||||||||       ||||||
Db        436 CAAAACCAGAACCTGAGGATGACTGGGTTCGGTCTTTACAACCAGAACATTAACCCAGAC 377

Qy         59 ---AsnGlnArgIlePheLeuAlaGlyArgProPhePheLeuAsnHis 73
                 ||||||||||||||:::||||||:::         :::||||||
Db        376 CATAACCAGAGGATCTTCGTGGCTGGGAAA---------ATCAACCAT 338

RESULT 2
US-09-731-375A-1
; Sequence 1, Application US/09731375A
; Publication No. US20030035810A1
; GENERAL INFORMATION:
;  APPLICANT: Caplan, Michael
;  TITLE OF INVENTION: Microbial Delivery System
;  FILE REFERENCE: 2002834-0100
;  CURRENT APPLICATION NUMBER: US/09/731,375A
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 60/195,035
;  PRIOR FILING DATE: 2000-04-06
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 2032
;   TYPE: DNA
;   ORGANISM: Arachis hypogaea
US-09-731-375A-1

Alignment Scores:
Pred. No.:              8.32e-13      Length:        2032
Score:                  156.50        Matches:       34
Percent Similarity:     66.67%        Conservative:  12
Best Local Similarity:  49.28%        Mismatches:    18
Query Match:            34.17%        Indels:        5
DB:                     10            Gaps:          2

US-10-691-590-1 (1-89) x US-09-731-375A-1 (1-2032)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              |||||||||:::||||||:::   :::||||||      |||      ::::::|||:::
Db       1484 CGGCGGGAAGAAGAGGAGGACGAAGACGAAGAAGAGGAGGGAAGTAACAGAGAGGTGCGT 1543

Qy         21 ---------AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaVal 37
                      |||   |||      |||||||||||:::   ||||||   |||   |||
Db       1544 AGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCTCATCCAGTA 1603

Qy         38 ThrPheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57
                        ||||||   :::      |||:::   :::   ||||||:::||||||:::
Db       1604 GCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATCAACGCTGAA 1657

Qy         58 AsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||   ||||||||||||||||||
Db       1658 AACAACCACAGAATCTTCCTTGCAGGT 1684

RESULT 3
US-10-228-806-1
; Sequence 1, Application US/10228806
; Publication No. US20030049237A1
; GENERAL INFORMATION:
;  APPLICANT: Bannon, et al.
;  TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reactions
;  TITLE OF INVENTION:  to Allergy
;  FILE REFERENCE: 2002834-0043
;  CURRENT APPLICATION NUMBER: US/10/228,806
;  CURRENT FILING DATE:  2002-08-26
;  NUMBER OF SEQ ID NOS: 81
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 2032
;   TYPE: DNA
;   ORGANISM: Arachis hypogaea
US-10-228-806-1

Alignment Scores:
Pred. No.:              8.32e-13      Length:        2032
Score:                  156.50        Matches:       34
Percent Similarity:     66.67%        Conservative:  12
Best Local Similarity:  49.28%        Mismatches:    18
Query Match:            34.17%        Indels:        5
DB:                     15            Gaps:          2

US-10-691-590-1 (1-89) x US-10-228-806-1 (1-2032)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              |||||||||:::||||||:::   :::||||||      |||      ::::::|||:::
Db       1484 CGGCGGGAAGAAGAGGAGGACGAAGACGAAGAAGAGGAGGGAAGTAACAGAGAGGTGCGT 1543

Qy         21 ---------AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaVal 37
                      |||   |||      |||||||||||:::   ||||||   |||   |||
Db       1544 AGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCTCATCCAGTA 1603

Qy         38 ThrPheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57
                        ||||||   :::      |||:::   :::   ||||||:::||||||:::
Db       1604 GCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATCAACGCTGAA 1657

Qy         58 AsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||   ||||||||||||||||||
Db       1658 AACAACCACAGAATCTTCCTTGCAGGT 1684

RESULT 4
US-10-100-303A-5
; Sequence 5, Application US/10100303A
; Publication No. US20030202980A1
; GENERAL INFORMATION:
;  APPLICANT: Caplan, et al.
;  TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reaction
;  TITLE OF INVENTION:  to Allergy
;  FILE REFERENCE: 2002834-0166
;  CURRENT APPLICATION NUMBER: US/10/100,303A
;  CURRENT FILING DATE:  2002-03-18
;  NUMBER OF SEQ ID NOS: 138
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 2032
;   TYPE: DNA
;   ORGANISM: Arachis hypogaea, Prot/Nucleo-Ara h 1
US-10-100-303A-5
```

```
Alignment Scores:
Pred. No.:              8.32e-13        Length:         2032
Score:                  156.50          Matches:        34
Percent Similarity:     66.67%          Conservative:   12
Best Local Similarity:  49.28%          Mismatches:     18
Query Match:            34.17%          Indels:         5
DB:                     18              Gaps:           2

US-10-691-590-1 (1-89) x US-10-100-303A-5 (1-2032)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              |||||||||:::|||||:::   :::||||||      |||      ::::::|||:::
Db       1484 CGGCGGGAAGAAGAGGAGGACGAAGACGAAGAAGAGGAGGGAAGTAACAGAGAGGTGCGT 1543

Qy         21 ---------AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaVal 37
                      |||   |||      |||||||||||:::   ||||||   |||   |||
Db       1544 AGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCTCATCCAGTA 1603

Qy         38 ThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57
                     ||||||   :::      |||:::   :::   ||||||:::||||||:::
Db       1604 GCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATCAACGCTGAA 1657

Qy         58 AsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||   ||||||||||||||||||
Db       1658 AACAACCACAGAATCTTCCTTGCAGGT 1684

RESULT 5
US-10-728-323-1
; Sequence 1, Application US/10728323
; Publication No. US20040208894A1
; GENERAL INFORMATION:
;  APPLICANT: Caplan, Michael J.
;  APPLICANT:  Bottomly H., Kim
;  APPLICANT:  Sosin B., Howard
;  APPLICANT:  Burks A., Wesley
;  APPLICANT:  Sampson A., Hugh
;  TITLE OF INVENTION: Microbial Delivery System
;  FILE REFERENCE: 2002834-0232
;  CURRENT APPLICATION NUMBER: US/10/728,323
;  CURRENT FILING DATE:  2003-12-04
;  PRIOR APPLICATION NUMBER: 60/195,035
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: 09/731,375
;  PRIOR FILING DATE: 2000-12-06
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 2032
;   TYPE: DNA
;   ORGANISM: Arachis hypogaea
US-10-728-323-1

Alignment Scores:
Pred. No.:              8.32e-13        Length:         2032
Score:                  156.50          Matches:        34
Percent Similarity:     66.67%          Conservative:   12
Best Local Similarity:  49.28%          Mismatches:     18
Query Match:            34.17%          Indels:         5
DB:                     21              Gaps:           2

US-10-691-590-1 (1-89) x US-10-728-323-1 (1-2032)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              |||||||||:::|||||:::   :::||||||      |||      ::::::|||:::
Db       1484 CGGCGGGAAGAAGAGGAGGACGAAGACGAAGAAGAGGAGGGAAGTAACAGAGAGGTGCGT 1543

Qy         21 ---------AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaVal 37
                      |||   |||      |||||||||||:::   ||||||   |||   |||
Db       1544 AGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCTCATCCAGTA 1603

Qy         38 ThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57
                     ||||||   :::      |||:::   :::   ||||||:::||||||:::
Db       1604 GCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATCAACGCTGAA 1657

Qy         58 AsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||   ||||||||||||||||||
Db       1658 AACAACCACAGAATCTTCCTTGCAGGT 1684

RESULT 6
US-10-728-051-1
; Sequence 1, Application US/10728051
; Publication No. US20040234548A1
; GENERAL INFORMATION:
;  APPLICANT: Caplan, Michael J.
;  APPLICANT:  Bottomly H., Kim
;  APPLICANT:  Sosin B., Howard
;  APPLICANT:  Burks A., Wesley
;  APPLICANT:  Sampson A., Hugh
;  TITLE OF INVENTION: Microbial Delivery System
;  FILE REFERENCE: 2002834-0222
;  CURRENT APPLICATION NUMBER: US/10/728,051
;  CURRENT FILING DATE:  2003-12-04
;  PRIOR APPLICATION NUMBER: 60/195,035
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: 09/731,375
;  PRIOR FILING DATE: 2000-12-06
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 2032
;   TYPE: DNA
;   ORGANISM: Arachis hypogaea
US-10-728-051-1

Alignment Scores:
Pred. No.:              8.32e-13        Length:         2032
Score:                  156.50          Matches:        34
Percent Similarity:     66.67%          Conservative:   12
Best Local Similarity:  49.28%          Mismatches:     18
Query Match:            34.17%          Indels:         5
DB:                     21              Gaps:           2

US-10-691-590-1 (1-89) x US-10-728-051-1 (1-2032)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              |||||||||:::|||||:::   :::||||||      |||      ::::::|||:::
Db       1484 CGGCGGGAAGAAGAGGAGGACGAAGACGAAGAAGAGGAGGGAAGTAACAGAGAGGTGCGT 1543

Qy         21 ---------AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaVal 37
                      |||   |||      |||||||||||:::   ||||||   |||   |||
Db       1544 AGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCTCATCCAGTA 1603

Qy         38 ThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57
                     ||||||   :::      |||:::   :::   ||||||:::||||||:::
Db       1604 GCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATCAACGCTGAA 1657

Qy         58 AsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||   ||||||||||||||||||
Db       1658 AACAACCACAGAATCTTCCTTGCAGGT 1684

RESULT 7
US-10-899-551-1
; Sequence 1, Application US/10899551
; Publication No. US20050063994A1
; GENERAL INFORMATION:
;  APPLICANT: Caplan, Michael J.
;  APPLICANT:  Burks, A. Wesley
;  APPLICANT:  Sampson, Hugh A.
;  APPLICANT:  Howard, Sosin B.
;  APPLICANT:  Bottomly, Kim H.
;  TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reaction to Allergy
;  FILE REFERENCE: 2002834-0233
;  CURRENT APPLICATION NUMBER: US/10/899,551
;  CURRENT FILING DATE:  2004-07-26
;  NUMBER OF SEQ ID NOS: 61
```

```
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 2032
;   TYPE: DNA
;   ORGANISM: species Arachis hypogea
US-10-899-551-1

Alignment Scores:
Pred. No.:                8.32e-13        Length:         2032
Score:                    156.50          Matches:        34
Percent Similarity:       66.67%          Conservative:   12
Best Local Similarity:    49.28%          Mismatches:     18
Query Match:              34.17%          Indels:         5
DB:                       22              Gaps:           2

US-10-691-590-1 (1-89) x US-10-899-551-1 (1-2032)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              |||||||||:::||||||:::   :::|||||||      |||      ::::::|||:::
Db       1484 CGGCGGGAAGAAGAGGAGGACGAAGACGAAGAAGAGGAGGGAAGTAACAGAGAGGTGCGT 1543

Qy         21 ---------AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaVal 37
                       |||   |||       ||||||||||||:::   ||||||   |||   |||
Db       1544 AGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCTCATCCAGTA 1603

Qy         38 ThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57
                     ||||||   :::      |||:::   :::   ||||||:::||||||:::
Db       1604 GCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATCAACGCTGAA 1657

Qy         58 AsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||   ||||||||||||||||||
Db       1658 AACAACCACAGAATCTTCCTTGCAGGT 1684

RESULT 8
US-10-958-324-5
; Sequence 5, Application US/10958324
; Publication No. US20050114924A1
; GENERAL INFORMATION:
;  APPLICANT: DODO, HORTENSE W.
;  APPLICANT:  ARNTZEN, CHARLES J.
;  APPLICANT:  KONAN, KOFFI N'DA
;  APPLICANT:  VIQUEZ, OLGA
;  TITLE OF INVENTION: DOWN-REGULATION AND SILENCING OF ALLERGEN GENES IN
;  TITLE OF INVENTION:  TRANSGENIC PEANUT SEEDS
;  FILE REFERENCE: 072121/0104
;  CURRENT APPLICATION NUMBER: US/10/958,324
;  CURRENT FILING DATE:  2004-10-06
;  PRIOR APPLICATION NUMBER: US/09/715,036
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: 60/167,255
;  PRIOR FILING DATE: 1999-11-19
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 2032
;   TYPE: DNA
;   ORGANISM: Arachis hypogaea
US-10-958-324-5

Alignment Scores:
Pred. No.:                8.32e-13        Length:         2032
Score:                    156.50          Matches:        34
Percent Similarity:       66.67%          Conservative:   12
Best Local Similarity:    49.28%          Mismatches:     18
Query Match:              34.17%          Indels:         5
DB:                       22              Gaps:           2

US-10-691-590-1 (1-89) x US-10-958-324-5 (1-2032)

Qy          1 ArgArgGluGlnGluGluGluSerGluGluGluThrPheGlyGluPheGlnGlnValLys 20
              |||||||||:::||||||:::   :::|||||||      |||      ::::::|||:::
Db       1484 CGGCGGGAAGAAGAGGAGGACGAAGACGAAGAAGAGGAGGGAAGTAACAGAGAGGTGCGT 1543

Qy         21 ---------AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaVal 37
                       |||   |||       ||||||||||||:::   ||||||   |||   |||
Db       1544 AGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCTCATCCAGTA 1603

Qy         38 ThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57
                     ||||||   :::      |||:::   :::   ||||||:::||||||:::
Db       1604 GCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATCAACGCTGAA 1657

Qy         58 AsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||   ||||||||||||||||||
Db       1658 AACAACCACAGAATCTTCCTTGCAGGT 1684

RESULT 9
US-10-021-323-2828/c
; Sequence 2828, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
;  APPLICANT: Deikman, Jill
;  APPLICANT:  Feng, Paul C.C.
;  APPLICANT:  Fincher, Karen L.
;  APPLICANT:  Ziegler, Todd E.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(52274)B
;  CURRENT APPLICATION NUMBER: US/10/021,323
;  CURRENT FILING DATE:  2001-12-12
;  PRIOR APPLICATION NUMBER: US 60/255, 619
;  PRIOR FILING DATE: 2000-12-14
;  NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 2828
;   LENGTH: 498
;   TYPE: DNA
;   ORGANISM: Gossypium hirsutum
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3825-018-Q1-N6-E12
US-10-021-323-2828

Alignment Scores:
Pred. No.:                2.18e-13        Length:         498
Score:                    154.50          Matches:        34
Percent Similarity:       66.20%          Conservative:   13
Best Local Similarity:    47.89%          Mismatches:     15
Query Match:              33.73%          Indels:         9
DB:                       20              Gaps:           2

US-10-691-590-1 (1-89) x US-10-021-323-2828 (1-498)

Qy         17 GlnGlnValLysAlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAla 36
              ::::::|||:::|||   |||||||||||||||:::||||||   ||||||||||||
Db        498 AAAAAAGTGAGAGCCCGGTTATCACCCGGTGACATATTTGTAGTCCCAGCAGGCCATCCA 439

Qy         37 ValThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeu---Asn 55
              |||||||||   ||||||:::::::::   |||              |||||||||   |||
Db        438 GTCACATTTGTTGCCTCCCAAAACAAAAACCTAGGATTGTTCGGGTTTGGACTTTACAAC 379

Qy         56 AlaGlnAsnAsnGlnArgIlePheLeuAlaGlyArgProPhePheLeuAsnHisLysGln 75
                 ||||||||||:::|||||||||:::   |||:::
Db        378 GGCCAAAACAACAAAAGGATTTTTGTGGTTGGGAAA---------------------- 343

Qy         76 AsnThrAsnValIleLysPheThrValLysAla 86
              |||||||||   ::::::      :::|||
Db        342 ACCACCAATGTTAGACAATGGGTTCGCCAGGCC 310

RESULT 10
US-10-437-963-43667
; Sequence 43667, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
```

```
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 43667
;   LENGTH: 2003
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_46801C.1
US-10-437-963-43667

Alignment Scores:
Pred. No.:              4.78e-12      Length:          2003
Score:                  151.50        Matches:         30
Percent Similarity:     60.53%        Conservative:    16
Best Local Similarity:  39.47%        Mismatches:      19
Query Match:            33.08%        Indels:          11
DB:                     20            Gaps:            1

US-10-691-590-1 (1-89) x US-10-437-963-43667 (1-2003)

Qy          2 ArgGluGlnGluGluGluSerGluGluGluThrPheGlyGlu----------------- 15
              ||||||:::|||||||||   ||||||:::     |||:::
Db       1253 CGTGAGGAGGAAGAGGAGGAGGAGGAGCAGCAGAAGGGGCAGGAGGAGGAAGAGGAAGAA 1312

Qy         16 --------------PheGlnGlnValLysAlaProLeuSerProGlyAspValPheVal 30
                           ::::::   ::::::|||   ||||||   |||   |||||||||
Db       1313 CAAGTCGGCCAGGGGTACGAGACGATCCGTGCGCGGCTGTCGCGCGGCACGGTGTTCGTC 1372

Qy         31 AlaProAlaGlyHisAlaValThrPhePheAlaSerLysAspGlnProLeuAsnAlaVal 50
                 |||:::||||||   :::         :::|||:::|||      |||      |||
Db       1373 GTCCCCTCCGGCCACCCGATCGTGGTGACGTCGTCCCGCGACAGCACCCTCCAGATCGTC 1432

Qy         51 AlaPheGlyLeuAsnAlaGlnAsnAsnGlnArgIlePheLeuAlaGly 66
                 |||   ::::::|||   ||||||:::|||::::::|||||||||
Db       1433 TGCTTCGACGTCCACGCCAACAACAACGAGAGGATGTACCTCGCCGGG 1480

RESULT 11
US-10-100-303A-6
; Sequence 6, Application US/10100303A
; Publication No. US20030202980A1
; GENERAL INFORMATION:
;  APPLICANT: Caplan, et al.
;  TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reaction
;  TITLE OF INVENTION:  to Allergy
;  FILE REFERENCE: 2002834-0166
;  CURRENT APPLICATION NUMBER: US/10/100,303A
;  CURRENT FILING DATE:  2002-03-18
;  NUMBER OF SEQ ID NOS: 138
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 1949
;   TYPE: DNA
;   ORGANISM: Arachis hypogaea, Prot/Nucleo-Ara h 1
US-10-100-303A-6

Alignment Scores:
Pred. No.:              5.48e-12      Length:          1949
Score:                  151.00        Matches:         36
Percent Similarity:     63.89%        Conservative:    10
Best Local Similarity:  50.00%        Mismatches:      18
Query Match:            32.97%        Indels:          8
DB:                     18            Gaps:            3

US-10-691-590-1 (1-89) x US-10-100-303A-6 (1-1949)

Qy          1 ArgArgGluGln---------GluGluGluSerGluGluGluThrPheGlyGluPheGln 17
                ||||||||||          |||||||| :::||||||      |||      :::
Db       1413 CGGCGGGAACAAGAGTGGGAAGAAGAGGAGGAAGATGAAGAAGAGGAGGGAAGTAACAGA 1472

Qy         18 GlnValLys---------AlaProLeuSerProGlyAspValPheValAlaProAlaGly 34
              :::|||:::          |||  |||      ||||||||||:::   ||||||
Db       1473 GAGGTGCGTAGGTACACAGCGAGGTTGAAGGAAGGCGATGTGTTCATCATGCCAGCAGCT 1532

Qy         35 HisAlaValThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeu 54
                |||   |||          ||||||   :::      |||:::   :::   ||||||:::
Db       1533 CATCCAGTAGCCATCAACGCTTCCTCCGAA------CTCCATCTGCTTGGCTTCGGTATC 1586

Qy         55 AsnAlaGlnAsnAsnGlnArgIlePheLeuAlaGly 66
              ||||||:::||||||   ||||||||||||||||||
Db       1587 AACGCTGAAAACAACCACAGAATCTTCCTTGCAGGT 1622

RESULT 12
US-10-425-114-7095
; Sequence 7095, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 7095
;   LENGTH: 1587
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 700617015_FLI
US-10-425-114-7095

Alignment Scores:
Pred. No.:              2.85e-11      Length:          1587
Score:                  145.50        Matches:         31
Percent Similarity:     57.75%        Conservative:    10
Best Local Similarity:  43.66%        Mismatches:      25
Query Match:            31.77%        Indels:          5
DB:                     19            Gaps:            1

US-10-691-590-1 (1-89) x US-10-425-114-7095 (1-1587)

Qy          1 ArgArgGluGlnGluGluGluSerGlu--------------GluGluThrPheGlyGlu 15
              ||||||   :::|||||||||   |||                ||||||
Db        812 AGGAGGAGCGAAGAAGAAGAAGAAGAATCGTCTGAGGAGCAGGAGGAGGCCGGGCAGGGG 871

Qy         16 PheGlnGlnValLysAlaProLeuSerProGlyAspValPheValAlaProAlaGlyHis 35
              :::      ::::::|||   ||||||||||||      ||||||   ||||||||||||
Db        872 TACCACACCATCCGGGCGCGGCTGTCACCGGGCACGGCGTTCGTGGTGCCCGCGGGCCAC 931

Qy         36 AlaValThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsn 55
                                 ||||||:::|||      |||      |||   |||   ::::::
Db        932 CCGTTCGTCGCGGTGGCGTCCCGGGACAGCAACCTCCAGATCGTGTGCTTCGAGGTCCAC 991

Qy         56 AlaGlnAsnAsnGlnArgIlePheLeuAlaGly 66
              |||      |||:::::::::|||||||||||
Db        992 GCCGACAGGAACGAGAAGGTGTTCCTGGCCGGC 1024

RESULT 13
US-10-425-114-23659
; Sequence 23659, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
```

```
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 23659
;   LENGTH: 1933
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3600-011-E12_FLI
US-10-425-114-23659

Alignment Scores:
Pred. No.:              3.79e-11        Length:         1933
Score:                  145.50          Matches:        31
Percent Similarity:     57.75%          Conservative:   10
Best Local Similarity:  43.66%          Mismatches:     25
Query Match:            31.77%          Indels:         5
DB:                     19              Gaps:           1

US-10-691-590-1 (1-89) x US-10-425-114-23659 (1-1933)

Qy          1 ArgArgGluGlnGluGluGluSerGlu--------------GluGluThrPheGlyGlu 15
              ||||||   :::|||||||||   |||                ||||||
Db       1218 AGGAGGAGCGAAGAAGAAGAAGAAGAATCGTCTGAGGAGCAGGAGGAGGCCGGGCAGGGG 1277

Qy         16 PheGlnGlnValLysAlaProLeuSerProGlyAspValPheValAlaProAlaGlyHis 35
              :::      ::::::|||   ||||||||||||      ||||||   ||||||||||||
Db       1278 TACCACACCATCCGGGCGCGGCTGTCACCGGGCACGGCGTTCGTGGTGCCCGCGGGCCAC 1337

Qy         36 AlaValThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsn 55
                        ||||||:::|||      |||      |||   |||   ::::::
Db       1338 CCGTTCGTCGCGGTGGCGTCCCGGGACAGCAACCTCCAGATCGTGTGCTTCGAGGTCCAC 1397

Qy         56 AlaGlnAsnAsnGlnArgIlePheLeuAlaGly 66
              |||      |||:::::::::|||||||||||
Db       1398 GCCGACAGGAACGAGAAGGTGTTCCTGGCCGGC 1430

RESULT 14
US-10-425-114-3415
; Sequence 3415, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 3415
;   LENGTH: 2001
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 700264357_FLI
US-10-425-114-3415

Alignment Scores:
Pred. No.:              3.99e-11        Length:         2001
Score:                  145.50          Matches:        31
Percent Similarity:     57.75%          Conservative:   10
Best Local Similarity:  43.66%          Mismatches:     25
Query Match:            31.77%          Indels:         5
DB:                     19              Gaps:           1

US-10-691-590-1 (1-89) x US-10-425-114-3415 (1-2001)

Qy          1 ArgArgGluGlnGluGluGluSerGlu--------------GluGluThrPheGlyGlu 15
              ||||||   :::|||||||||   |||                ||||||
Db       1223 AGGAGGAGCGAAGAAGAAGAAGAAGAATCGTCTGAGGAGCAGGAGGAGGCCGGGCAGGGG 1282

Qy         16 PheGlnGlnValLysAlaProLeuSerProGlyAspValPheValAlaProAlaGlyHis 35
              :::      ::::::|||   ||||||||||||      ||||||   ||||||||||||
Db       1283 TACCACACCATCCGGGCGCGGCTGTCACCGGGCACGGCGTTCGTGGTGCCCGCGGGCCAC 1342

Qy         36 AlaValThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsn 55
                        ||||||:::|||      |||      |||   |||   ::::::
Db       1343 CCGTTCGTCGCGGTGGCGTCCCGGGACAGCAACCTCCAGATCGTGTGCTTCGAGGTCCAC 1402

Qy         56 AlaGlnAsnAsnGlnArgIlePheLeuAlaGly 66
              |||      |||:::::::::|||||||||||
Db       1403 GCCGACAGGAACGAGAAGGTGTTCCTGGCCGGC 1435

RESULT 15
US-10-425-115-117618
; Sequence 117618, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 117618
;   LENGTH: 2926
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(2926)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_38757C.1
US-10-425-115-117618

Alignment Scores:
Pred. No.:              6.93e-11        Length:         2926
Score:                  145.50          Matches:        31
Percent Similarity:     57.75%          Conservative:   10
Best Local Similarity:  43.66%          Mismatches:     25
Query Match:            31.77%          Indels:         5
DB:                     21              Gaps:           1

US-10-691-590-1 (1-89) x US-10-425-115-117618 (1-2926)

Qy          1 ArgArgGluGlnGluGluGluSerGlu--------------GluGluThrPheGlyGlu 15
              ||||||   :::|||||||||   |||                ||||||
Db       1795 AGGAGGAGCGAAGAAGAAGAAGAAGAATCGTCTGAGGAGCAGGAGGAGGCCGGGCAGGGG 1854

Qy         16 PheGlnGlnValLysAlaProLeuSerProGlyAspValPheValAlaProAlaGlyHis 35
              :::      ::::::|||   ||||||||||||      ||||||   ||||||||||||
Db       1855 TACCACACCATCCGGGCGCGGCTGTCACCGGGCACGGCGTTCGTGGTGCCCGCGGGCCAC 1914

Qy         36 AlaValThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsn 55
                        ||||||:::|||      |||      |||   |||   ::::::
Db       1915 CCGTTCGTCGCGGTGGCGTCCCGGGACAGCAACCTCCAGATCGTGTGCTTCGAGGTCCAC 1974

Qy         56 AlaGlnAsnAsnGlnArgIlePheLeuAlaGly 66
              |||      |||:::::::::|||||||||||
```

Db 1975 GCCGACAGGAACGAGAAGGTGTTCCTGGCCGGC 2007

Search completed: November 5, 2005, 06:20:50
Job time : 907.835 secs